THE ANNALS [illegible]

MATHEMATICAL STATISTICS

[illegible]

[illegible] OF MATHEMATICAL STATISTICS

Contents

Vol. [illegible], No. [illegible]

THE [illegible]NALS OF [illegible]HEMA[illegible] STATISTICS

E. L[illegible]

A[illegible] E[illegible]

[illegible] BLA[illegible] J. L. HODGES, JR. [illegible]LLIAM G. MA[illegible]
[illegible]ANIE[illegible] WASSILY HOEFFDING [illegible] WOLFOWITZ

WITH [illegible] OF

[illegible] BIRNBAUM D. A. D[illegible] G. E. NOETHER
[illegible] BOSE J. L. DOOB [illegible] PERSAKO[illegible]
[illegible] T. E. HARRIS [illegible] E. ROBBINS
[illegible] PAUL G. HOEL [illegible] J. SAVAGE
[illegible] J. KIEFER [illegible]ARLES M. STEIN
[illegible] WILLIAM [illegible] [illegible] WEISS
[illegible] [illegible] [illegible] A. WOODBURY

[illegible] by the In[illegible] of Mathematical Statistics in March, [illegible] September and [illegible] at [illegible], Maryland

INSTITUTE OF MATH[illegible]ATICAL STATISTICS

Treasurer's Office Department of [illegible], Stanford University, [illegible], Cal. A. H. Bowker, T[illegible]
These addresses [illegible] be used for communications [illegible] scriptions, back [illegible], payment of dues, etc.

Secretary's Office Department of Sta[illegible] of North Carolina, Ch[illegible] N. C.
George E. Nicholson [illegible]
These addresses should [illegible] ship, changes of address [illegible] effective for a given issue the [illegible] reported to the Secretary on or before the 10th of the month [illegible] the month of issue.

From August 1, 195[illegible]

Editorial Office The RAND Corporation, [illegible] Main St., Santa Monica, Cal.
T. E. Harris, Editor
Manuscripts should [illegible] to this address. [illegible] should be typewritten, double-spaced with wide margins, [illegible] original copy should be submitted (preferably with one additional [illegible]). Ditto and mimeographed papers are acceptable if they are com[illegible] legible and printed on paper that will take corrections. Footnotes [illegible] be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and [illegible] should be professionally drawn on plain white paper or tracing [illegible] in black India ink twice the size they are to be printed. Authors [illegible] to keep in mind typographical difficulties of complicated mathematical formulae. Authors will ordinarily receive only galley [illegible]. Fifty reprints with[illegible] covers will be furnished free. Additional re[illegible] covers furnished at cost.

Subscription Prices Current issues are [illegible] calendar [illegible] the U. S. and Canada [illegible] Single issues [illegible] Back numbers are [illegible] volume [illegible] $200 for the [illegible] volumes, $10 per [illegible] volume [illegible] at the same time as Volume I through XX[illegible]

C[illegible] P[illegible]
WAVERLY P[illegible], BALTIMORE, M[illegible]LAND, U. S. A.

MEYER ABRAHAM GIRSHICK 1908–1955

By David Blackwell and Albert H. Bowker

Meyer Abraham Girshick, a Fellow of the Institute and its president in 1952, died in the Palo Alto Hospital on March 2, 1955 at the age of 46. He was born in a small Russian village and came to New York City at the age of 15 years in 1923. The principal of the elementary school he attended in New York was Angelo Patri who took a strong interest in the boy and helped him get into Columbia College in 1929. In 1932 he married Mary Knabel. In 1934 he entered graduate school at Columbia University to work with Professor Harold Hotelling who arranged a stipend from a Carnegie Foundation grant.

Girshick left Columbia in 1937 to begin a very distinguished career in government service. For the next ten years he held positions in several government and government sponsored agencies including the Bureau of Home Economics and the Bureau of Agricultural Economics in the Department of Agriculture, the Statistical Research Group at Columbia University, the Bureau of the Census, and the Rand Corporation in Santa Monica. He joined the staff of Stanford University as Professor of Statistics in 1948. He is survived by his wife Mary and their daughter Paula.

After he left Columbia in 1937, he undertook a pioneer study [3] of body measurements of 147,000 American children for the purpose of helping manufacturers of clothing develop an improved system of sizing garments. At the same time he began a series of evening courses at the Department of Agriculture graduate school. Through these courses he attracted many research workers to the field and played an important part in promoting the use of sound statistical methods in the federal government.

He moved from the Bureau of Home Economics, to become principal statistician in the Bureau of Agricultural Economics in 1939, a position which he left to join the Statistical Research Group at Columbia University. This period of activity at SRG had a decisive influence on his career. While there, he participated in the development of sequential analysis and wrote his two most important papers [10], [11] in this area. During this period he became acquainted with and immediately recognized the importance of the new and more sophisticated decision theory models for statistical problems being developed by Wald. From about 1946 most of his work was explicitly formulated in terms of loss functions and other decision theory concepts. This interest was reinforced by his work in games at the Rand Corporation, and a major portion of his time in recent years was spent in an effort to clarify and extend the basic results of decision theory [23]. Girshick soon found himself surrounded by students, junior colleagues, and others in the University who sought his advice, counsel, and encouragement in their work.

At the time of the Korean war, Girshick organized a military research group at Stanford with the sponsorship of the office of Naval Research. His leadership

of the applied statistics group at Stanford has long been considered a model of University participation in military research, and through his efforts many University scientists have been able to contribute directly to difficult problems in theoretical statistics of military interest. His intellectual leadership in both the Statistics Department and projects, and enthusiastic interest in scholarly work were major factors in the growth of Statistics at Stanford. Most of the work produced by the Statistics Department represents his ideas or his spirit.

At the time of his death, he was exploring the role of invariance in statistical problems, an interest reflected earlier in [18]. This work was continued actively at Stanford University and became one of the major themes of research in the growing Statistics Department.

Girshick was notable for his receptivity to new concepts (sequential analysis, decision theory, game theory, invariance), his tremendous energy and drive, the wealth of new ideas and conjectures he produced, and his persistent and usually successful efforts to get others to work in directions he considered fruitful. His influence in statistics was at least as much through the impact he had on all who came in contact with him as through his own writings and will be felt for a long time.

BIBLIOGRAPHY

[1] "Principle Components." *J. Amer. Stat. Assn.*, Vol. 31. 1936.

[2] "On the Sampling Theory of Roots and Determinantal Equations." *Ann. Math. Stat.*, Vol. 10. 1939.

[3] *Body Measurements of American Boys and Girls* (co-author with Ruth O'Brien and Eleanor Hunt). U. S. Dept. of Agr. Misc. Pub. No. 366. 1941.

[4] "The Distribution of the Ellipticity Statistic Le when the Hypothesis is False." *Journal of Terrestrial Magnetism and Atmospheric Electricity*. 1941.

[5] "Note on the Distribution of Roots of a Polynomial with Random Complex Coefficients." *Ann. Math. Stat.*, Vol. 8. 1942.

[6] "The Application of the Theory of Linear Hypothesis to the Coefficient of Elasticity of Demand." *J. Amer. Stat. Assn.*, Vol. 37. 1942.

[7] "Some Extensions of the Wishart Distribution" (co-author with T. W. Anderson). *Ann. Math. Stat.*., Vol. 15. 1944.

[8] *Sequential Analysis of Statistical Data, Applications* (co-author). Columbia Univ. Press. 1945.

[9] "Unbiased Estimates for Certain Binomial Sampling Plans with Applications" (co-author with F. Mosteller and L. J. Savage). *Ann. Math. Stat.*, Vol. 17. 1946.

[10] "Contributions to the Theory of Sequential Analysis I." *Ann. Math. Stat.*, Vol. 17. 1946.

[11] "Contributions to the Theory of Sequential Analysis II." *Ann. Math. Stat.*, Vol. 17. 1946.

[12] "Functions of Sequences of Independent Chance Vectors with Applications to the Problem of the Random Walk in N Dimensions (co-author with D. Blackwell). *Ann. Math. Stat.*, Vol. 17. 1946.

[13] "A Lower Bound for the Variance of Some Unbiased Sequential Estimates" (co-author with D. Blackwell). *Ann. Math. Stat.*, Vol. 17. 1947.

[14] "Statistical Analysis of the Demand for Food," (co-author with Trygve Haavelmo). *Econometrica*. Vol. 15. 1947.

[15] "Bayes and Minimax Solutions of Sequential Decision Problems" (with K. J. Arrow and David Blackwell). *Econometrica*, Vol. 17. 1949.

[16] "The Prediction of Social and Technological Events" (with A. Kaplan and A. L. Skogstad). *Public Opinion Quarterly*, Vol. 14. 1950.

[17] "Model Construction in the Social Sciences—an Expository Discussion of Measurement and Prediction (with Daniel Lerner). *Public Opinion Quarterly*, Vol. 14. 1950-51.

[18] "Bayes and Minimax Estimates for Quadratic Loss Function" (with L. J. Savage). *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.

[19] "A Bayes Approach to a Quality Control Model" (with Herman Rubin). *Ann. Math. Stat.*, Vol. 23. 1952.

[20] "Estimates of Bounded Relative Error in Particle Counting" (co-author with Herman Rubin and Rosedith Sitgreaves). *Ann. Math. Stat.*, Vol. 26. 1952.

[21] "Optimal Invariant Statistical Decision Procedures" (co-author with D. Blackwell and H. Rubin), to be published. 1953.

[22] "Optimal Invariant Sequential Decision Procedures" (co-author with Herman Rubin), to be published. 1953.

[23] *Theory of Games and Statistical Decisions* (co-author with D. Blackwell). John Wiley and Sons, Inc. 1954.

[24] "An Elementary Survey of Statistical Decision Theory." *Review of Educational Research*. December 1954.

THE TREATMENT OF TIES IN SOME NONPARAMETRIC TESTS[1]

By Joseph Putter

University of California, Berkeley

1. Introduction. Most of nonparametric testing theory is usually presented under the assumption that all the samples involved are drawn from continuous distributions, and that tied observations can therefore be ignored or treated in any convenient way, without affecting the performance characteristic of the test. In practice, however, this assumption is not a realistic one, and the distributions involved are in general to be regarded as discontinuous, either because of intrinsic reasons (integer-valued or otherwise discrete random variables) or because of limitations on the precision of measurements. Therefore, usually, ties will occur with positive probability, and the way they are treated does affect the performance characteristic of the test. The problem of ties has therefore to be considered, in particular with a view to preserving the nonparametric character of the test, and to making sure of setting it up on the desired level of significance.

The usual practice in attacking the problem has been to consider the conditional distributions of the statistics concerned given that the number of observations in each tied group is a fixed constant. This, however, was never explicitly made clear, and these conditional distributions, as well as their variances and other characteristics, are referred to as distributions (or variances, etc.) "when ties are present." In this category belong Kendall's work on ties in rank correlation theory, and Kruskal's theorem concerning a generalized Wilcoxon test (see Section 8).

In this paper, we attack the problem from the standpoint of the ties being random variables. Our main concern is the comparison between the "randomized" and the "nonrandomized" way of treating the ties. In Sections 3 and 4 we consider the one-sided sign test, and show that randomization reduces both the exact power and the asymptotic efficiency of the test. In Sections 5–8 we consider the Wilcoxon test. For small samples the nonrandomized treatment of ties presents practical difficulties, but the asymptotic (large sample) problem can be handled. Again, it is shown that randomization results in reduced efficiency.

2. Notation and theorems used. We shall use the notation $\mathfrak{N}(a, b)$ for normal random variables (with mean a and variance b), and $\mathfrak{B}(n, p)$ for binomials. The symbol $\xrightarrow{P}$ will denote convergence in probability, and $\xrightarrow{L}$ convergence in law (convergence of distributions).

Received March 16, 1954.

[1] Work supported in part by the Office of Naval Research, and submitted in partial satisfaction of the requirements for the degree of Ph.D. at the University of California, Berkeley.

To compare the asymptotic performances of two consistent tests, we shall use Pitman's concept of asymptotic relative efficiency. The concept is presented in Pitman's lecture notes and also by Noether [11]. In particular, we shall use the following theorem.

THEOREM A (Pitman, as quoted in Noether [11], pp. 241–242). *Let H be a hypothesis specifying the value θ_0 of a population parameter θ, and A the one-sided alternative $\theta > \theta_0$. Let $\{\tau_{in}\}$, $i = 1, 2$; $n = 1, 2, \cdots$, be two sequences of tests of H against A, on the same level of significance α. Let τ_{in} consist of rejecting H when $\bar{S}_{in} > k_{in}$, where $\bar{S}_{in}$ are statistics and k_{in} appropriate constants. Let $\psi_{in}(\theta)$ and $\sigma_{in}(\theta)$ be functions such that $\psi'_{in}(\theta)$ exists in the neighborhood of θ_0, and let the following conditions be satisfied as $n \to \infty$:*

$$\frac{\psi'_{in}(\theta_n)}{\psi'_{in}(\theta_0)} \to 1, \qquad \theta_n = \theta_0 + \frac{a}{n^{1/2}}, \quad a \text{ a positive constant}; \tag{2.1}$$

$$\frac{\sigma_{in}(\theta_n)}{\sigma_{in}(\theta_0)} \to 1; \tag{2.2}$$

$$\frac{H_i(n)}{n^{1/2}} \to c_i, \qquad H_i(n) = \frac{\psi'_{in}(\theta_0)}{\sigma_{in}(\theta_0)}, \quad c_i \text{ a positive constant}; \tag{2.3}$$

and either

$$\frac{\bar{S}_{in} - \psi_{in}(\theta)}{\sigma_{in}(\theta)} \xrightarrow{L} \mathfrak{N}(0, 1) \tag{2.4}$$

uniformly in θ in the neighborhood of θ_0, or

$$\frac{\bar{S}_{in} - \psi_{in}(\theta_n)}{\sigma_{in}(\theta_n)} \xrightarrow{L} \mathfrak{N}(0, 1). \tag{2.5}$$

Then the asymptotic relative efficiency of $\{\tau_{2n}\}$ with respect to $\{\tau_{1n}\}$ is $\lim_{n\to\infty} H_2^2(n)/H_1^2(n)$.

(Noether defines $\psi_{in}(\theta) = E(\bar{S}_{in} \mid \theta)$ and $\sigma_{in}(\theta) = \sigma(\bar{S}_{in} \mid \theta)$, but it is easily seen from Pitman's proof of the theorem that this specification is not necessary.)

To handle the uniform convergence required in condition (2.4), we shall use the following theorem.

THEOREM B (*Parzen* [12], p. 35). *A necessary and sufficient condition for a sequence of distributions $F_n = F_n^{(\theta)}$ to converge to a distribution F uniformly in θ is that*

$$f_n^{(\theta)}(t) \to f(t) \text{ uniformly in } \theta, \tag{2.6}$$

where $f_n^{(\theta)}$ and f denote the respective characteristic functions. The convergence (2.6) is then jointly uniform in θ and t for every finite t-interval.

(Theorem B is a particular case of Parzen's Theorem 7c.)

THE SIGN TEST

3. Randomized and nonrandomized test. Let $Z_1, \cdots, Z_n$ be independent and identically distributed random variables. Denote the number of positive

Z_k's by N_+, of negative Z_k's by N_-, and of zeros among the Z_k's by N_0. The sign test consists of rejecting the hypothesis

$$H: \quad P(Z_k > 0) = P(Z_k < 0),$$

against the alternative

$$A: \quad P(Z_k > 0) > P(Z_k < 0),$$

say, whenever N_+ is too large.

In practice Z_k frequently is of the form $X_k - Y_k$, where X_k and Y_k are independent. If the distribution functions of X_k and Y_k are continuous, then $P(Z_k = 0) = 0$. In this case, under the hypothesis, N_+ is $\mathfrak{B}(n, \frac{1}{2})$, which gives us the cutoff point.

In the general (discontinuous) case, denote

$$P(Z_k > O \mid H) = p_+, \qquad P(Z_k = O \mid H) = p_0,$$

$$P(Z_k > O \mid A) = q_+, \qquad P(Z_k = O \mid A) = q_0, \qquad P(Z_k < O \mid A) = q_-.$$

Consider the conditional distribution of N_+ given that $N_0 = n_0$. Under H,

$$P(N_+ = x \mid n_0) = p_H(x) = \binom{n - n_0}{x} (\tfrac{1}{2})^{n-n_0};$$

under A,

$$P(N_+ = x \mid n_0) = p_A(x) = \binom{n - n_0}{x} \left(\frac{q_-}{1 - q_0}\right)^{n-n_0} \left(\frac{q_+}{q_-}\right)^x,$$

$x = 0, 1, \cdots, n - n_0$. Thence

$$\frac{p_A(x)}{p_H(x)} = c(n_0) \left(\frac{q_+}{q_-}\right)^x,$$

which is a strictly increasing function of x. Therefore, by the Neyman-Pearson lemma, the unique most powerful test based on N_+ and N_0 is given by

$$N_+ > k(N_0), \tag{3.1}$$

where the cutoff point $k(n_0)$ is, of course, the one corresponding to $\mathfrak{B}(n - n_0, \frac{1}{2})$.

It is obvious that $k(N_0)$ is not a linear function of N_0. Thence the test (3.1) does not coincide with the test

$$N_+ + \tfrac{1}{2}N_0 > k, \tag{3.2}$$

which was proposed, e.g., by Dixon and Mood [3]. In fact, the distribution of $N_+ + \frac{1}{2}N_0$ under H depends on the unknown parameter p_0, so that the cutoff point k cannot be well defined. The usual practice seems to be to take for k the cutoff point corresponding to $\mathfrak{B}(n, \frac{1}{2})$. This, as was shown by Hemelrijk [4], results in lowering the level of significance and consequently also the power of

the test. However, the difficulty caused by the dependence of k on p_0 can be obviated when asymptotic properties are considered, and we shall return to the matter in the next section.

The test (3.1), which amounts to "omitting the ties from the observations" and which was suggested, e.g., by Dixon and Massey [2], is, as we have seen, the unique most powerful test based on N_+ and N_0. However, another customary procedure is one based on "randomization": after observing the Z_k's, we perform N_0 independent random experiments, assigning each of the N_0 zeros among the Z_k's a positive or negative sign with equal probabilities ($= \frac{1}{2}$). We thus get, say, N_+^r additional positives. The random variable $N_+^R = N_+ + N_+^r$ is, under H, $\mathcal{B}(n, \frac{1}{2})$, and we can apply the test

$$N_+^R > k \tag{3.3}$$

without worrying about the unknown p_0.

Consider, again, the conditional situation given that $N_0 = n_0$. Denote by $p(y)$ the frequency distribution of $\mathcal{B}(n_0, \frac{1}{2})$. The joint (conditional) frequency distribution of N_+ and N_+^r is $p_H(x)p(y)$ under H, and $p_A(x)p(y)$ under A. The ratio of the two expressions is $p_A(x)/p_H(x)$, so that (3.1) is also the unique most powerful test based on N_+, N_0, *and* N_+^r. We have thus proved the following theorem.

THEOREM 1. *The nonrandomized test* (3.1) *is uniformly more powerful* (*against the one-sided alternative* A) *than the randomized test* (3.3).

As a numerical example, we give in Table I the powers of the two tests for $n = 10$, against the alternative $q_+/q_- = 2$. Since the power of either test depends on q_0, we tabulate the conditional power given $N_0 = n_0$, for all values of n_0. The tests are considered on the .05 level. To keep this level exact (and to get a valid comparison between the tests), we modify the tests in the usual way. For example, the test (3.3) is now formulated as follows: Reject H with probability 1 if $N_+^R > k$; reject H with probability φ if $N_+^R = k$; accept H otherwise. In our particular case, $k = 8$ and $\varphi = .893$.

TABLE I

n_0	0	1	2	3	4	5	6	7	8	9	10
Power of randomized test (3.3)	.278	.241	.208	.177	.150	.127	.106	.089	.074	.061	.050
Power of nonrandomized test (3.1)	.278	.244	.232	.216	.184	.171	.158	.119	.088	.067	.050

In particular, against the alternative $q_0 = q_- = \frac{1}{4}$, $q_+ = \frac{1}{2}$, the power of (3.3) is .195, while that of (3.1) is .221.

4. Asymptotic properties. For large sample sizes n, it is convenient to use the normal approximation to the binomial, and we shall now compare the performances of the randomized and nonrandomized tests when this approximation is used.

For the randomized test statistic N_+^R we have, under H,

$$\frac{2N_+^R - n}{n^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1), \tag{4.1}$$

which gives us the usual normal approximation to (3.3). Under A, N_+^R is

$$\mathfrak{B}(n, q_+ + \tfrac{1}{2}q_0),$$

and hence

$$\frac{N_+^R - n(q_+ + \frac{1}{2}q_0)}{[n(q_+ + \frac{1}{2}q_0)(q_- + \frac{1}{2}q_0)]^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1). \tag{4.2}$$

It is easily seen from (4.1) and (4.2) that the test (3.3) is consistent against the one-sided alternative A.

It is more difficult to derive the normal approximation to (3.1). Since in any case *the* normal approximation of a test is not an obviously definable concept, we shall derive a nonrandomized asymptotic test by starting from (3.2). The joint distribution of N_+, N_0, and N_- is trinomial; hence the test statistic

$$N'_+ = N_+ + \tfrac{1}{2}N_0$$

is asymptotically normal. More precisely, under H,

$$\frac{2N'_+ - n}{[n(1 - p_0)]^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1).$$

Since $N_0/n \xrightarrow{P} p_0$, we have

$$\frac{2N'_+ - n}{(n - N_0)^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1), \tag{4.3}$$

which gives us an asymptotic test independent of p_0. Under A,

$$\frac{N'_+ - n(q_+ + \frac{1}{2}q_0)}{(n[(q_+ + \frac{1}{2}q_0)(q_- + \frac{1}{2}q_0) - \frac{1}{4}q_0])^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1), \tag{4.4}$$

and, again, the nonrandomized test corresponding to (4.3) is consistent against A.

We now compare the asymptotic performances of the two tests in terms of Pitman's concept of asymptotic relative efficiency. In the notation of Theorem A, put

$$\theta = q_+ + \tfrac{1}{2}q_0, \qquad \theta_0 = \tfrac{1}{2}.$$

THEOREM 2. *Let $\{A_\theta, \theta > \frac{1}{2}\}$ be a family of alternatives for which $q_0 = p_0$. Then the asymptotic relative efficiency of the randomized test (3.3) with respect to the (nonrandomized) test based on the statistic*

$$T_n = \frac{2N'_+ - n}{(n - N_0)^{1/2}}$$

is $1 - p_0$.

Proof. Put

$$\bar{S}_{1n} = T_n(1 - p_0)^{1/2}, \quad \bar{S}_{2n} = N_+^R,$$
$$\psi_{1n}(\theta) = (2\theta - 1)n^{1/2}, \quad \psi_{2n}(\theta) = n\theta,$$
$$\sigma_{1n}(\theta) = (4\theta(1 - \theta) - p_0)^{1/2}, \quad \sigma_{2n}(\theta) = (n\theta(1 - \theta))^{1/2}.$$

Conditions (2.1)–(2.3) obviously hold, and we proceed to verify (2.4) and/or (2.5).

For $i = 2$, the convergence (2.4) holds by (4.2). From the usual proof of binomial convergence to the normal, it is easily seen that the corresponding convergence of characteristic functions is uniform in θ in the neighborhood of θ_0, and hence, by Theorem B, so is (4.2), and condition (2.4) holds. For $i = 1$, we have

$$\frac{\bar{S}_{1n} - \psi_{1n}(\theta)}{\sigma_{1n}(\theta)} = \frac{2(N_+ - \theta n)}{(n[4\theta(1 - \theta) - p_0])^{1/2}} \left(\frac{n(1 - p_0)}{n - N_0}\right)^{1/2}$$
$$+ (2\theta - 1)\left(\frac{n(1 - p_0)}{4\theta(1 - \theta) - p_0}\right)^{1/2}\left(\left(\frac{n}{n - N_0}\right)^{1/2} - \frac{1}{(1 - p_0)^{1/2}}\right).$$

Now,

$$\frac{2(N_+ - \theta n)}{(n[4\theta(1 - \theta) - p_0])^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1)$$

by (4.4), and this convergence, as before, can easily be shown to be uniform in θ in the neighborhood of θ_0. We have

$$\left(\frac{n(1 - p_0)}{n - N_0}\right)^{1/2} \xrightarrow{P} 1 \quad \text{and} \quad \left(\frac{n}{n - N_0}\right)^{1/2} - \frac{1}{(1 - p_0)^{1/2}} \xrightarrow{P} 0$$

independently of θ, and

$$(2\theta_n - 1)\left(\frac{n(1 - p_0)}{4\theta_n(1 - \theta_n) - p_0}\right)^{1/2} \to 2a.$$

Hence condition (2.5) holds. Our result now follows from Theorem A.

THE WILCOXON TEST

5. Notation and known results. We shall use the following notation in connection with the Wilcoxon test. $(X_1, \cdots, X_n)$ is a sample of n independent observations from a distribution $F(z)$, and $(Y_1, \cdots, Y_m)$ is a sample of m independent observations from a distribution $G(z)$. If all the $m + n$ observations in the pooled sample are different, we rank them in ascending order of magnitude, assigning the rank 1 to the smallest observation. We denote by S_{nm} the sum of the ranks assigned to the X's. The Wilcoxon test of the hypothesis $F = G$ consists of rejecting the hypothesis when S_{nm} is too large.

The mean, variance, and asymptotic distribution of S_{nm}, in the case when F and G are continuous (and therefore the probability of getting two or more equal observations is 0), are known, and are summarized below.

When $F = G$, every possible ordering of the pooled sample occurs with the same probability $1/(n + m)!$, and the distribution of S_{nm} can be derived from this fact alone. We shall denote any statistic with this probability distribution by S^0_{nm}. From Mann-Whitney [10] we have

$$(5.1)\qquad ES^0_{nm} = \frac{n(n + m + 1)}{2} = \mu_{nm}, \text{ say;}$$

$$(5.2)\qquad \sigma^2(S^0_{nm}) = \frac{nm(n + m + 1)}{12} = \sigma^2_{nm}, \text{ say;}$$

$$(5.3)\qquad T^0_{nm} = \frac{S^0_{nm} - \mu_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, 1) \quad \text{as } \frac{1}{n} + \frac{1}{m} \to 0.$$

In general, when F and G are any two (continuous) distributions, we have, from Mann-Whitney [10],

$$(5.4)\qquad ES_{nm} = \mu_{nm} + nm\theta,$$

$$(5.5)\qquad \theta = \theta(F, G) = P(X_1 > Y_1) - \tfrac{1}{2} = \int_{-\infty}^{\infty} G(z)\, dF(z) - \tfrac{1}{2},$$

$$(5.6)\qquad \begin{aligned}\sigma^2(S_{nm}) = \sigma^2_{nm} + nm[(\theta - \lambda_1)(n - 1) \\ + (\theta - \lambda_2)(m - 1) - \theta^2(n + m - 1)],\end{aligned}$$

$$(5.7)\qquad \lambda_1 = \lambda_1(F, G) = \tfrac{1}{3} - \int_{-\infty}^{\infty} F^2(z)\, dG(z),$$

$$(5.8)\qquad \lambda_2 = \lambda_2(F, G) = \tfrac{1}{3} - \int_{-\infty}^{\infty} [1 - G(z)]^2\, dF(z).$$

When $n \to \infty$ while $m/n = c$ is held constant, we have, from Lehmann [9],

$$(5.9)\qquad \frac{S_{nm} - ES_{nm}}{\sigma(S_{nm})} \xrightarrow{L} \mathfrak{N}(0, 1).$$

(That $\sigma(S_{nm})$ is the correct norming factor can be seen from Hoeffding [6], Theorem 5.2.)

For the case of discontinuous F and G, which we shall consider in the following sections, we adopt the following notation. We assume the common discontinuities of F and G (which are the only ones that matter) to be finite in number, and denote them by ξ_k, $k = 1, \cdots, K$. Their locations are not assumed known, and are irrelevant to our considerations. We define

$p_k = P(X_1 = \xi_k), \qquad q_k = P(Y_1 = \xi_k);$

$U_k =$ the number of X's which are equal to ξ_k;

$V_k =$ the number of Y's which are equal to ξ_k;

$W_k = U_k + V_k$;

$$U = (U_1, \cdots, U_K), \qquad V = (V_1, \cdots, V_K), \qquad W = (W_1, \cdots, W_K).$$

We shall write $\sum_k$ for $\sum_{k=1}^{K}$.

6. The treatment of ties. When F and G are continuous, the probability of getting tied (equal) observations is 0, so that this event may be ignored. In the discontinuous case, however, ties occur with positive probability, and when they do occur, the pooled sample can no longer be uniquely ordered. The problem arises, therefore, of how the Wilcoxon test is to be defined in such a case.

An obvious solution to the problem, proposed by many writers on the subject, is, again, "randomization": each group of equal observations is ordered at random, giving every possible ordering (within the group) the same probability. This results in an ordering of the pooled sample, and the sum of the ranks of the X's can now be defined. The only difference from the continuous case is that this new random variable is defined over a different sample space, because its value depends not only on the observed X's and Y's but also on the outcome of the randomization procedure. We shall denote this sum of the "randomized" ranks of the X's also by S_{nm}.

Again, if $F = G$, every possible (randomized) ordering of the pooled sample is equally probable, and hence S_{nm} is distributed as S^0_{nm}. The Wilcoxon test can therefore be applied using S_{nm}, with the same cutoff point as in the continuous case. The main objection to this procedure seems to be that the outcome of the test (rejection or acceptance of the hypothesis) is thus made to depend not only on the observations but also on an additional, and more or less irrelevant, random experiment. We are thus led to look for a test which is

(i) distribution-free under the hypothesis;
(ii) dependent on the observations only; and
(iii) as close as possible to the original Wilcoxon test.

We leave the precise meaning of this last requirement unspecified for the moment, and shall elaborate the point later on.

For the remainder of this section, we shall need to consider only the case when $F = G$, and it will be convenient to assume that F is purely discontinuous. In this case, the ordering of the pooled sample is given by the nonzero component of the two vectors U and V, as long as the observations alone are considered. Hence any rank (order) statistic which depends on the observations only can be expressed in terms of U and V. Requirement (ii) means, therefore, that the rejection region R of the test will be a region in the $2K$-dimensional sample space of the random vector $(U_1, \cdots, U_K, V_1, \cdots, V_K)$.

In this sample space, the vector W is a sufficient statistic for the vector parameter $(p_1, \cdots, p_K)$, i.e., the conditional probability

$$P(u \mid w) = P(U = u, V = w - u \mid W = w)$$
$$= \frac{n!}{u_1! \cdots u_K!} \cdot \frac{m!}{(w_1 - u_1)! \cdots (w_K - u_K)!} \bigg/ \frac{(m+n)!}{w_1! \cdots w_K!}$$

is independent of the p_k's. Hence, if the size α of R, that is,

$$P(R) = \sum_w P(W = w)\, P(R \mid W = w)$$
$$= \sum_w \frac{(m+n)!}{w_1! \cdots w_K!} p_1^{w_1} \cdots p_K^{w_K} \sum_{(u, w-u) \in R} P(u \mid w),$$

is to be independent of the p_k's (requirement (i)), we must have $P(R \mid W = w) = \alpha$ for every w, which is the usual condition on distribution-free tests when a sufficient statistic with a complete family is involved. But since for every fixed w we have only a finite set of probabilities $P(u \mid w)$, and these sets vary with w, it will in general be impossible to find a region R with exact size α. However, this difficulty can be obviated, e.g., by considering regions which include some sample points not definitely but with certain given probabilities. Thus it appears that some random element outside the observations is unavoidable, unless we do not insist on the exact size α. But in practice this consideration is unimportant, because one is usually quite content to stop just short of the given size α.

Suppose that various regions R of the required type and of exact size α (produced by the above, or any other, device) are available. Denote by R_0 the rejection region $[S_{nm} > a]$, of the same size α, given by the "randomized" Wilcoxon test. Then R_0 is defined in a different sample space, which can be described as the result of splitting each point of the (u, v)-space into several points corresponding to the possible outcomes of the randomization procedure. We shall view the sets R as sets in this "extended" sample space, too.

We have $P(R) = P(R_0) = \alpha$, or $P(R \cap \bar{R}_0) = P(\bar{R} \cap R_0)$, where the notation $\bar{A}$ stands for the complement of A. Now, one possible interpretation of requirement (iii) above is to choose R so as to minimize $P(R \cap \bar{R}_0)$. This may be justified as follows. Suppose F is really continuous, and the ties occur only because of insufficient precision of measurement. The randomized test is, in a sense, approximately equivalent to the (Wilcoxon) test which we would use if our measurements were precise, because the effect of the randomization procedure is similar to the effect of replacing each discontinuity by an interval of uniform distribution (cf. Section 7). It is therefore reasonable to try to minimize the probability of getting a result (rejection or acceptance of the hypothesis) different from the result of the randomized test. But this probability, when the hypothesis is true, is $P(R \cap \bar{R}_0) + P(\bar{R} \cap R_0) = 2P(R \cap \bar{R}_0)$.

We thus want to minimize $P(R \cap [S_{nm} \leqq a])$, which will be achieved if we minimize

$$P(R \cap [S_{nm} \leqq a] \mid W = w) = \sum_{(u, w-u) \epsilon R} P(u \mid w)\, P(S \leqq a \mid U = u, V = w - u)$$

for every w. This is to be done under the condition

$$\sum_{(u, w-u) \epsilon R} P(u \mid w) = P(R \mid W = w) = \alpha. \tag{6.1}$$

In a manner analogous to the proof of the Neyman-Pearson lemma, it is easily seen that the "optimum" region is obtained by the following procedure. For every vector w, we order all the possible vectors $(u, v) = (u, w - u)$ by the magnitude of $P(S \leqq a \mid U = u, V = w - u)$. We take that vector $(u, w - u)$ for which this probability is smallest, then that vector for which it is the next smallest, etc., until the (conditional) size α, as in (6.1), is reached. Doing this for all w, we get the desired R.

Unfortunately, the tabulations required for this test are much too extensive. We can approximate the test if, instead of rejecting the hypothesis when $P(S_{nm} \leq a \mid U = u, V = w - u)$ is too small, we reject it when $E(S_{nm} \mid U = u, V = w - u)$ is too large. The two tests will probably not differ too much.

The statistic

$$S'_{nm} = E_r(S_{nm} \mid U, V), \tag{6.2}$$

where E_r denotes expected value under randomization, is the sum of the midranks of the X's, where the midrank of an observation is defined as the mean rank of all the observations equal to that observation, or, more precisely,

$$\text{midrank } (X) = \frac{N_1 + N_2 + 1}{2},$$

where N_1 is the number of observations smaller than X, and N_2 is the number of observations (including X) not larger than X. The statistic S'_{nm} has been proposed by many writers as a test statistic to replace S_{nm} when ties are present. However, by the preceding considerations, the cutoff point has to depend on W, and the tabulation involved is prohibitive. A few cases have been tabulated in [13], but they can merely serve as an indication of the task involved in more exhaustive tabulation.

Kruskal [7] derived the conditional asymptotic distribution of S'_{nm} given fixed $W = w(n, m)$ which fulfill a certain convergence condition (cf. Section 8) for the case $F = G$. In the next two sections we shall derive the (unconditional) asymptotic distribution of S'_{nm} in general, and discuss some consequences.

7. The asymptotic distribution of S'_{nm}. We now drop the assumption that F and G are purely discontinuous. Consider the conditional distribution of S_{nm} given a fixed pooled sample of X's and Y's. For this fixed sample, let $U = u$, $V = v$. Denote by r the sum of the ranks of those X's which are not equal to any ξ_k (and which are therefore, with probability 1, untied), and by r_k the number of those observations (X's and Y's) which are smaller than ξ_k.

Under the randomization procedure which generates S_{nm}, those observations which are equal to ξ_k are assigned the ranks $r_k + 1, r_k + 2, \cdots, r_k + u_k + v_k$ at random, with every ordering equally probable. Hence the sum of the ranks of those X's which are equal to ξ_k is $u_k r_k + S^0_{u_k, v_k}$, and

$$S_{nm} = r + \textstyle\sum_k (u_k r_k + S^0_{u_k, v_k}).$$

Therefore, by (6.2) and (5.1),

$$S'_{nm} = r + \textstyle\sum_k (u_k r_k + \mu_{u_k, v_k}),$$

$$S_{nm} = S'_{nm} + \textstyle\sum_k (S^0_{u_k, v_k} - \mu_{u_k, v_k}),$$

where the $K + 1$ terms on the right are (conditionally) independent. Since this

holds for every fixed sample, we can write

$$S_{nm} = S'_{nm} + \sum_k (S^0_{U_k, V_k} - \mu_{U_k, V_k}), \tag{7.1}$$

where the terms on the right are conditionally independent given U and V.

Obviously, we have

$$ES'_{nm} = EE_r(S_{nm} \mid U, V) = ES_{nm}.$$

To calculate $\sigma^2(S_{nm})$, we note that, by (7.1),

$$\begin{aligned}\sigma^2(S_{nm}) &= E(S_{nm} - ES_{nm})^2 \\ &= EE_r[(S_{nm} - ES_{nm})^2 \mid U, V] \\ &= EE_r\{[S'_{nm} - ES_{nm} + \textstyle\sum_k (S^0_{U_k, V_k} - \mu_{U_k, V_k})]^2 \mid U, V\} \\ &= \sigma^2(S'_{nm}) + \tfrac{1}{12} \textstyle\sum_k EU_kV_k(U_k + V_k + 1) \\ &= \sigma^2(S'_{nm}) + \frac{nm}{12} \sum_k p_k q_k[(n-1)p_k + (m-1)q_k + 3],\end{aligned}$$

or

$$\sigma^2(S'_{nm}) = \sigma^2(S_{nm}) - \frac{nm}{12} \sum_k p_k q_k[(n-1)p_k + (m-1)q_k + 3]. \tag{7.2}$$

In particular, when $F = G$,

$$\sigma^2(S'_{nm}) = \sigma^2_{nm} - \frac{nm}{12} \sum_k p_k^2[(n + m - 2)p_k + 3]. \tag{7.3}$$

Of some interest, when $F = G$, is also the conditional variance $\sigma^2(S'_{nm} \mid W)$. Since the conditional distribution of S_{nm} given W is still that of S^0_{nm}, this variance can be computed in a manner similar to the preceding argument, giving (when $F = G$)

$$\sigma^2(S'_{nm} \mid W) = \sigma^2_{nm} - \frac{nm}{12(n+m)(n+m-1)} \sum_k W_k(W_k^2 - 1). \tag{7.4}$$

This is the variance given by Kruskal [7], and in a more cumbersome form by Hemelrijk [5].

Since ES_{nm} and $\sigma^2(S_{nm})$ as given by (5.4) and (5.6) refer to the continuous case, we shall touch on the modifications required for discontinuous F and G. By Lemma 5.1 of Lehmann [9], there exist two continuous distributions F^* and G^* under which the distribution of S_{nm} is the same as under F and G. These continuous distributions are obtained, essentially, by replacing the discontinuities by intervals of uniform distribution. We define

$$\begin{aligned}\theta^* &= \theta^*(F, G) = \theta(F^*, G^*), \\ \lambda_j^* &= \lambda_j^*(F, G) = \lambda_j(F^*, G^*), \qquad j = 1, 2,\end{aligned}$$

referring to the definitions (5.5), (5.7), and (5.8). From (5.4) and (5.6), we now have

$$ES_{nm} = \mu_{nm} + \theta^* nm, \tag{7.5}$$

$$\begin{aligned}\sigma^2(S_{nm}) = \sigma_{nm}^2 + nm[(\theta^* - \lambda_1^*)(n-1) \\ + (\theta^* - \lambda_2^*)(m-1) - \theta^{*2}(n+m-1)].\end{aligned} \tag{7.6}$$

For later use, we compute θ^* in terms of F and G. Denote by B the real line with the points ξ_k excluded. We have

$$\begin{aligned}\theta^* + \tfrac{1}{2} &= \int_{-\infty}^{\infty} G^*(z)\, dF^*(z) = \int_B G(z)\, dF(z) + \sum_k \int_0^1 [G(\xi_k - 0) + tq_k] p_k\, dt \\ &= \int_B G(z)\, dF(z) + \sum_k p_k G(\xi_k - 0) + \tfrac{1}{2} \sum_k p_k q_k = P(X_1 > Y_1) \\ &\qquad + \tfrac{1}{2} P(X_1 = Y_1),\end{aligned}$$

or

$$\theta^* = P(X_1 > Y_1) + \tfrac{1}{2} P(X_1 = Y_1) - \tfrac{1}{2}. \tag{7.7}$$

We now give a theorem connecting the asymptotic distribution of S'_{nm} with that of S_{nm}. Note that the symbol σ_{U_k,V_k}^2 will stand for $\frac{1}{12} U_k V_k (U_k + V_k + 1)$, as in (5.2), and will have nothing to do with the variances of U_k and V_k. This refers, of course, to all the symbols with U_k and V_k as subscripts.

THEOREM 3. *If, for a pair of distributions* (F, G), *and possibly under some restrictions concerning the relation between* n *and* m, *we have*

$$\frac{S_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, b^2), \tag{7.8}$$

$$\frac{\sigma_{U_k,V_k}^2}{\sigma_{nm}^2} \xrightarrow{P} b_k^2 \tag{7.9}$$

as $1/n + 1/m \to 0$, *then, under the same conditions,*

$$\frac{S'_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, \bar{b}^2),$$

where

$$\bar{b}^2 = b^2 - \textstyle\sum_k b_k^2.$$

PROOF. Subtracting ES_{nm} from both sides of (7.1) and dividing by σ_{nm}, we have

$$\frac{S_{nm} - ES_{nm}}{\sigma_{nm}} = \frac{S'_{nm} - ES_{nm}}{\sigma_{nm}} + \sum_k \frac{\sigma_{U_k,V_k}}{\sigma_{nm}} T_{U_k,V_k}^0, \tag{7.10}$$

where T^0 is defined by (5.3). The U_k and V_k are $\mathcal{B}(n, p_k)$ and $\mathcal{B}(m, q_k)$, respectively.

Let $d > 0$ be a fixed number which we shall specify later, and define

$$R_{nm} = \left[\left| \frac{U_k}{n} - p_k \right| < d, \left| \frac{V_k}{m} - q_k \right| < d, \left| \frac{\sigma_{U_k, V_k}}{\sigma_{nm}} - b_k \right| < d, \text{ all } k \right].$$

We have, by (7.9),

$$\text{(7.11)} \qquad P(R_{nm}) \to 1 \quad \text{as} \quad 1/n + 1/m \to 0.$$

Define

$h(t) \quad = e^{-t^2/2};$

$h_{nm}(t) =$ the characteristic function of T^0_{nm};

$f_{nm}(t) =$ the characteristic function of $\dfrac{S_{nm} - ES_{nm}}{\sigma_{nm}}$;

$f^{uv}_{nm}(t) =$ the conditional characteristic function of $\dfrac{S_{nm} - ES_{nm}}{\sigma_{nm}}$ given $U = u$, $V = v$;

$g_{nm}(t) =$ the characteristic function of $\dfrac{S'_{nm} - ES_{nm}}{\sigma_{nm}}$;

$g^{uv}_{nm}(t) =$ the conditional characteristic function of $\dfrac{S'_{nm} - ES_{nm}}{\sigma_{nm}}$ given $U = u$, $V = v$;

$A \quad = \prod_k h(tb_k), \qquad a_k = \dfrac{\sigma_{u_k, v_k}}{\sigma_{nm}}.$

All integrals will be taken in the (u, v)-space, with respect to the probability measure in that space.

By (7.10), we have

$$g^{uv}_{nm}(t) = \frac{f^{uv}_{nm}(t)}{\prod_k h_{u_k, v_k}(ta_k)},$$

or

$$g^{uv}_{nm}(t) - h(t\bar{b}) = \frac{1}{A} [f^{uv}_{nm}(t) - h(tb)]$$
$$+ \frac{1}{A} g^{uv}_{nm}(t) \left[\prod_k h(tb_k) - \prod_k h_{u_k, v_k}(ta_k) \right].$$

Hence

$$|g_{nm}(t) - h(t\bar{b})| = \left| \int [g^{uv}_{nm}(t) - h(t\bar{b})] \right|$$
$$\leqq \int_{\bar{R}_{nm}} |g^{uv}_{nm}(t) - h(t\bar{b})| + \frac{1}{A} \left| \int_{R_{nm}} [f^{uv}_{nm}(t) - h(tb)] \right|$$
$$+ \frac{1}{A} \int_{R_{nm}} \left| \prod_k h(tb_k) - \prod_k h_{u_k, v_k}(ta_k) \right|.$$

Using the definition of R_{nm}, the property (7.11), the condition (7.8), and the fact that, by (5.3), $h_{nm}(t) \to h(t)$, each of the three expressions involved can be shown to converge to 0 as $1/n + 1/m \to 0$, and the theorem is proved.

8. Consequences of Theorem 3. Since the asymptotic distribution of S_{nm} is known, Theorem 3 enables us to investigate the asymptotic behavior of S'_{nm} and to compare the tests based on the two statistics.

THEOREM 4. *If $F = G$, then*

$$\frac{S'_{nm} - \mu_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, 1 - \sum_k p_k^3), \text{ as } \frac{1}{n} + \frac{1}{m} \to 0.$$

Therefore, if $s_{nm} = s_{nm}(U, V)$ is any sequence of positive statistics satisfying

$$\frac{s_{nm}^2}{\sigma_{nm}^2} \xrightarrow{P} 1 - \sum_k p_k^3,$$

then

$$T'_{nm} = \frac{S'_{nm} - \mu_{nm}}{s_{nm}} \xrightarrow{L} \mathfrak{N}(0, 1) \text{ as } \frac{1}{n} + \frac{1}{m} \to 0.$$

PROOF. By (5.9), we know that (7.8) holds with $b = 1$. We also have

$$\frac{\sigma^2_{U_k, V_k}}{\sigma^2_{nm}} = \frac{U_k V_k (W_k + 1)}{nm(n + m + 1)}$$

$$= \frac{U_k}{n} \frac{V_k}{m} \left(\frac{W_k}{n + m} + \frac{n + m - W_k}{(n + m)(n + m + 1)} \right) \xrightarrow{P} p_k^3$$

Hence the theorem follows from Theorem 3.

Theorem 4 gives us test statistics whose asymptotic distributions, under the hypothesis $F = G$, are independent of F, and which can therefore be used to obtain asymptotically distribution-free tests. The rejection region of such a test will be $[T'_{nm} > a]$, where a is given by

$$(2\pi)^{-1/2} \int_a^\infty e^{-t^2/2} \, dt = \alpha.$$

We shall refer to tests of this type as "the nonrandomized tests," and to the Wilcoxon test, based on the randomized S_{nm}, as "the randomized test."

Convenient choices for the norming factor s_{nm} are given, e.g., by

$$s_{nm}^2 = \sigma_{nm}^2 - \tfrac{1}{12} \textstyle\sum_k U_k V_k (W_k + 1), \tag{8.1}$$

or

$$s_{nm}^2 = \sigma_{nm}^2 - \frac{nm}{12(n + m)(n + m - 1)} \sum_k W_k (W_k^2 - 1). \tag{8.2}$$

The norming factor given by (8.2), which is, by (7.4), the conditional standard deviation $\sigma(S'_{nm} \mid W)$ under the hypothesis, was suggested by Kruskal and Wallis [8]. In this case, Kruskal [7] proved that the conditional distribution of

T'_{nm} (given W) tends to $\mathfrak{N}(0, 1)$ if the W's are fixed vectors such that s_{nm}/σ_{nm} converges to a positive limit.

We now turn to the case $F \neq G$ (the alternative of the test). In the continuous case, it has been shown by Mann-Whitney [10], van Dantzig [1], and Lehmann [9] that if m/n is held constant, then the Wilcoxon test is consistent (i.e., its power tends to 1 as $n \to \infty$) against all alternatives under which $P(X_1 > Y_1) > \frac{1}{2}$. We proceed to derive the analogous consistency property for the discontinuous case.

THEOREM 5. *Let $m/n = c$ be fixed, and*

$$P(X_1 > Y_1) + \tfrac{1}{2}P(X_1 = Y_1) > \tfrac{1}{2}. \tag{8.3}$$

Then the randomized test is consistent. If, moreover, the norming factor s_{nm} satisfies

$$\frac{s_{nm}}{\sigma_{nm}} \xrightarrow{P} h > 0, \tag{8.4}$$

then the nonrandomized test is also consistent.

REMARK. The condition (8.4) is always satisfied if s_{nm} is defined by either (8.1) or (8.2). For (8.1), we have

$$h^2 = 1 - \frac{1}{1 + c} \sum_k p_k q_k (p_k + c q_k),$$

and for (8.2),

$$h^2 = 1 - \frac{1}{(1 + c)^3} \sum_k (p_k + c q_k)^3,$$

and both quantities are positive, unless F and G are both degenerate and identical, which is obviously impossible under (8.3).

PROOF OF THEOREM. By (5.9) and Lemma 5.1 of [9], we have

$$\frac{S_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, b^2), \tag{8.5}$$

where $b = \lim_{n \to \infty} \sigma(S_{nm})/\sigma_{nm}$ is, by (7.6), a function of c and of the parameters θ^*, λ_1^*, and λ_2^*. The rejection region of the randomized test is $[T_{nm} > a]$, where $T_{nm} = (S_{nm} - \mu_{nm})/\sigma_{nm}$. But, by (7.5), we have

$$T_{nm} = \frac{S_{nm} - ES_{nm}}{\sigma_{nm}} - \frac{\theta^* nm}{\sigma_{nm}},$$

and, by (5.2), $nm/\sigma_{nm} \to \infty$ as $n \to \infty$. Hence, by (7.7), the randomized test is consistent.

Also, from (8.5) it follows, by Theorem 3, that

$$\frac{S'_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, b^2),$$

where

$$\bar{b}^2 = b^2 - \frac{1}{1+c} \sum_k p_k q_k (p_k + c q_k).$$

Hence

$$\frac{S'_{nm} - ES_{nm}}{s_{nm}} \xrightarrow{L} \mathfrak{N}\left(0, \frac{\bar{b}^2}{\bar{h}^2}\right),$$

and the consistency of the nonrandomized test follows by the same argument as above, which completes the proof of the theorem.

9. Asymptotic efficiency. We shall now compare the randomized and the nonrandomized tests in terms of Pitman's concept of asymptotic relative efficiency. We shall restrict ourselves to the case of purely discontinuous distributions. Under a host of conditions (necessary to insure that the conditions of Theorem A are satisfied), it will be shown that the nonrandomized test is asymptotically more efficient than the randomized test, and that its asymptotic efficiency does not depend on the choice of the norming factor s_{nm}. The parameter θ will be $\theta^*(F, G) = P(X_1 > Y_1) + \frac{1}{2}P(X_1 = Y_1) - \frac{1}{2}$, and hence $\theta_0 = 0$.

LEMMA. *Let Z_i $(i = 1, 2, \cdots, r)$ be $\mathfrak{B}(na_i, p_i)$, and put $\bar{Z}_i = Z_i - na_i p_i$. Then*

$$\prod_{i=1}^{r} Z_i = n^{r-1}\left[n + \sum_{i=1}^{r} \frac{\bar{Z}_i}{a_i p_i}\right] \prod_{i=1}^{r} a_i p_i + o_p(n^{r-1/2}).$$

Here the notation $f_n = o_p(g_n)$ stands, as usual, for $f_n/g_n \xrightarrow{P} 0$. The proof of the lemma consists of expanding the product $\prod Z_i = \prod(\bar{Z}_i + na_i p_i)$ and noting that $\bar{Z}_i/n^{1/2}$ converges in law (to a normal).

THEOREM 6. *Let $m/n = c$ be fixed, and F be a purely discontinuous distribution. Let $\{G_\theta, 0 \leqq \theta \leqq \theta_1\}$ be a family of purely discontinuous distributions having the same discontinuities ξ_k as F, with jumps $q_k(\theta)$. Let $\{G_\theta\}$ have the following properties:*

(1) $G_0 = F$;

(2) $q_k(\theta) > q > 0$;

(3) $\theta^*(F, G_\theta) = \theta$;

(4) *the convergence* $(S_{nm} - ES_{nm})/\sigma_{nm} \xrightarrow{L} \mathfrak{N}(0, b^2(\theta))$, *given by* (8.5), *is uniform in* θ;

(5) *the functions* $q_k(\theta)$ *are continuous at* $\theta = 0$.

Let $s_{nm} = s_{nm}(U, V)$ be continuous functions of U and V, having, under (F, G_θ), finite variances and satisfying the following conditions:

(6) $s_{nm}/n^{1/2} = \sum_k \alpha_k(\theta)\bar{U}_k + \sum_k \beta_k(\theta)\bar{V}_k + \gamma(\theta)n + o_p(n^{1/2})$, *where*
$\bar{U}_k = U_k - np_k$, $\bar{V}_k - mq_k(\theta)$;

(7) $\gamma^2(0) = (c(1+c)/12)(1 - \sum_k p_k^3)$:

(8) $\gamma(\theta)$ *is differentiable, and* $\gamma'(\theta)$ *is continuous at* $\theta = 0$;

(9) *at least one of the* $2K$ *inequalities* $c\theta\alpha_k(\theta) \neq \gamma(\theta)\bar{\alpha}_k(\theta)$, $\quad c\theta\beta_k(\theta) \neq \gamma(\theta)\bar{\beta}_k$ *holds, where (arranging the* ξ_k*'s so that* $\xi_k < \xi_{k+1}$*)*

$$\bar{\alpha}_k(\theta) = 1 + c[\textstyle\sum_{j<k} q_j(\theta) + \frac{1}{2}q_k(\theta)], \qquad \bar{\beta}_k = \textstyle\sum_{j>k} p_j + \frac{1}{2}p_k .$$

Under these conditions, the asymptotic relative efficiency of the randomized test with respect to the nonrandomized test is $1 - \sum_k p_k^3$.

Remarks. (i) Conditions (6) and (9) are satisfied if s_{nm} is given either by (8.1) or by (8.2). For (8.1), for example, using the lemma in this section, we have

$$\frac{12}{c}\frac{s_{nm}^2}{n^3} = 1 + c - \sum_k p_k q_k(\theta)[p_k + cq_k(\theta)] - \sum_k q_k(\theta)[2p_k + cq_k(\theta)]\,\frac{\bar{U}_k}{n}$$
$$- \sum_k p_k[p_k + 2cq_k(\theta)]\,\frac{\bar{V}_k}{m} + o_p(n^{-1/2}),$$

and using the Taylor expansion for $(a + bx)^{1/2}$ we get (6). The same method works for (8.2).

(ii) Condition (7) is necessary, by Theorem 4, to make s_{nm} an admissible norming factor. It is satisfied for the choices (8.1) and (8.2) if

$$\left[\frac{d}{d\theta}\Big(1 - \sum_k q_k^3(\theta)\Big)\right]_{\theta=0} = 0,$$

which is analogous to the condition $q_0 = p_0$ in Theorem 2.

Proof of Theorem. In terms of Theorem A, put

$$\bar{S}_{1n} = T'_{nm}, \qquad \bar{S}_{2n} = S_{nm} - \mu_{nm},$$
$$\psi_{1n}(\theta) = \frac{c\theta}{\gamma(\theta)}\, n^{1/2}, \qquad \psi_{2n}(\theta) = \theta nm,$$
$$\sigma_{1n}(\theta) = \frac{c(1 + c)}{12n^3\gamma^2(\theta)}\, \sigma([S'_{nm} - \psi_{1n}(\theta)s_{nm}] \mid \theta),$$
$$\sigma_{2n}(\theta) = \sigma(S_{nm} \mid \theta).$$

We have then

$$H_1^2(n) = \frac{12cn}{(1 + c)(1 - \sum_k p_k^3)}, \qquad H_2^2(n) = \frac{n^2m^2}{\sigma_{nm}^2}.$$

The verification of (2.1)–(2.3) is routine, and we proceed to verify (2.4). The convergences involved are all uniform in θ; except for the one required by condition (4), this follows from condition (2) and Theorem B. (All the usual binomial convergences, when put in terms of characteristic functions, are seen to be uniform as long as the probability parameters are bounded away from 0 and 1.)

We have

$$\frac{\bar{S}_{2n} - \psi_{2n}(\theta)}{\sigma_{2n}(\theta)} = \frac{S_{nm} - ES_{nm}}{\sigma(S_{nm})},$$

which, by (5.9), verifies (2.4) for $i = 2$. Also, we have

$$\frac{S_{1n} - \psi_{1n}(\theta)}{\sigma_{1n}(\theta)} = \frac{S'_{nm} - \mu_{nm} - \psi_{1n}(\theta)s_{nm}}{\sigma_{1n}(\theta)s_{nm}}.$$

Arranging the ξ_k's in ascending order of magnitude, we have

$$S'_{nm} = \sum_k U_k\left(\sum_{j<k} W_j + \frac{W_k + 1}{2}\right).$$

It follows, by the lemma in this section, that

$$\frac{S'_{nm}}{n} = \sum_k \bar{\alpha}_k(\theta)\bar{U}_k + \sum_k \bar{\beta}_k\bar{V}_k + n\bar{\gamma}(\theta) + o_p(n^{1/2}),$$

where

$$\bar{\gamma}(\theta) = \textstyle\sum_k p_k\{\sum_{j<k}[p_j + cq_j(\theta)] + \frac{1}{2}[p_k + cq_k(\theta)]\}.$$

Hence

$$\frac{1}{n}[S'_{nm} - \mu_{nm} - \psi_{1n}(\theta)s_{nm}] = \sum_k\left[\bar{\alpha}_k(\theta) - \frac{c\theta}{\gamma(\theta)}\alpha_k(\theta)\right]\bar{U}_k$$
$$+ \sum_k\left[\bar{\beta}_k - \frac{c\theta}{\gamma(\theta)}\beta_k(\theta)\right]\bar{V}_k + o_p(n^{1/2}).$$

Because of (9), this expression is asymptotically normal, and it is easily shown that (2.4) holds for $i = 1$. Hence, by Theorem A, the asymptotic relative efficiency of the randomized test with respect to the nonrandomized test is

$$\lim_{n\to\infty}\frac{H_2^2(n)}{H_1^2(n)} = 1 - \sum_k p_k^3,$$

and the theorem is proved.

10. Acknowledgment. The author wishes to thank Professor E. L. Lehmann for his suggestion of the problem and for his helpful interest in the work.

REFERENCES

[1] D. van Dantzig, "On the consistency and the power of Wilcoxon's two-sample test," *Nederl. Akad. Wetensch., Proc.*, Vol. 54 (1951), pp. 1–8.

[2] W. J. Dixon and F. J. Massey, Jr., *An introduction to statistical analysis*, McGraw-Hill Book Co., 1951, p. 248.

[3] W. J. Dixon and A. M. Mood, "The statistical sign test," *J. Amer. Stat. Assn.*, Vol. 41 (1946), pp. 557–566.

[4] J. Hemelrijk, "A theorem on the sign test when ties are present," *Nederl. Akad. Wetensch., Proc.*, Vol. 55 (1952), pp. 322–326.

[5] J. Hemelrijk, "Note on Wilcoxon's two sample test when ties are present," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 133–135.

[6] W. Hoeffding, "A class of statistics with asymptotically normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293–325.

[7] W. H. Kruskal, "A nonparametric test for the several-sample problem," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 525–540.
[8] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 583–621.
[9] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.
[10] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 50–60.
[11] G. E. Noether, "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 231–246.
[12] E. Parzen, "On uniform convergence of families of sequences of random variables," *Univ. California Publ. Stat.*, Vol. 2 (1954), pp. 23–54.
[13] *Probability Tables for the Wilcoxon Test When There are Ties*, National Bureau of Standards Report No. 1859, Washington, D. C.

ON A CLASS OF DECISION PROCEDURES FOR RANKING MEANS OF NORMAL POPULATIONS[1]

By K. C. Seal

University of North Carolina

Summary. An infinite class of decision rules having several desirable properties is suggested for choosing a group of populations from a given set of normal populations which should contain the population with the largest mean. The problem of selecting one member from this infinite class of rules has also been studied.

1. Introduction. In recent years it has been recognized ([1], [3], [4], [5], [9], [10], [11]) that the conventional test of homogeneity, such as the F-test in the analysis of variance for testing the equality of several population means, does not supply all the information that the experimenter seeks. In many practical situations it is unrealistic to assume that the population means of several essentially different populations will be equal. A sufficiently large sample will thus enable the experimenter to detect this difference at any preassigned level. In most cases what the experimenter actually wants is a decision procedure which would tell him which population or populations possess a desired characteristic. For example, the experimenter may be interested in determining the population with the largest mean, from a set of normal populations. Alternatively he may desire to select from a given number of populations a group containing the population having the largest mean.

Suppose there are $n + 1$ normal populations $N(\mu_i, \sigma_1^2)$, $i = 0, 1, 2, \cdots, n$, with unknown means and a common but unknown variance and that k random observations $x_{i\alpha}$ $(i = 0, 1, \cdots, n;\ \alpha = 1, 2, \cdots, k)$ from each of the $n + 1$ normal populations are given, where $x_{i\alpha}$ is one of the k observations from the ith population. Under our assumptions the $n + 1$ sample means

$$x_i = \sum_{\alpha=1}^{k} x_{i\alpha} / k$$

will obey $N(\mu_i, \sigma_1^2 / k)$, $i = 0, 1, \cdots, n$, and an estimate

$$s_1^2 = \sum_{i=0}^{n} \sum_{\alpha=1}^{k} (x_{i\alpha} - x_i)^2 / (k - 1)(n + 1)$$

of σ_1^2 can be obtained which is independent of the sample means x_i, $i = 0, 1, \cdots, n$. We may, therefore, assume for mathematical convenience that just one random observation x_i from each of $n + 1$ normal populations $N(\mu_i, \sigma^2)$, $i =$

Received July 6, 1954.

[1] This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

$0, 1, \cdots, n$, is given, where $\sigma^2 = \sigma_1^2 / k$ is estimated by

$$s^2 = s_1^2 / k = \sum_{i=0}^{n} \sum_{\alpha=1}^{k} (x_{i\alpha} - x_i)^2 / [k(k-1)(n+1)]$$

and this is assumed to be known. Clearly this estimate s^2 of σ^2 is stochastically independent of the given observations x_i, $i = 0, 1, \cdots, n$. It is desired to choose a group of populations from the above $n + 1$ populations, with the help of some decision rule which ensures that the least upper bound of the probability of not including in the group the population with the largest mean is α $(0 < \alpha < 1)$, whatever may be the unknown μ_i's. Subject to this fundamental requirement we would like the rule to possess other desirable properties such as:

(a) The property of unbiasedness, i.e., the probability of rejecting any population not having the largest mean is not less than the probability of rejecting the population having the largest mean. (Analogy of this property to the property of unbiasedness in the theory of testing of hypothesis should be noticed.)

(b) The property of gradation, i.e., corresponding to any α $(0 < \alpha < 1)$, there exists a constant $\mu_{0\alpha}$ such that the chance of retaining the population with mean μ_0 in the group is greater or less than α, according as μ_0 is greater or less than $\mu_{0\alpha}$. The constant $\mu_{0\alpha}$ will in general depend on the decision rule as well as the unknown means of the remaining n populations, and the common variance σ^2.

An infinite class $\mathcal{C}$ of decision rules satisfying the fundamental requirement, together with the properties (a) and (b) is given in Section 2.1. Certain interesting properties of this class are studied in Section 3. The question of choosing one member from this infinite class having further desirable properties has been studied in Section 4.

2. Class $\mathcal{C}$ of decision procedures.

2.1. Let y_i, $i = 0, 1, \cdots, n$, be $n + 1$ random observations from $N(0, \sigma^2)$ and let $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$ be n ranked observations among $y_1, \cdots, y_n$. The $y_{(i)}$'s will then define another set of random variables $Y_{(i)}$, $i = 1, \cdots, n$. It is assumed $y_i \neq y_j$, $i \neq j$, since the set of points $(y_0, y_1, \cdots, y_n)$ in $(n + 1)$-dimensional Euclidean space where $y_i \neq y_j$, $i \neq j$ will be obtained with probability 1. Let $t_\alpha(c_1, \cdots, c_n)$ $(c_i \geqq 0, i = 1, \cdots, n, \sum_1^n c_i = 1)$ denote the *upper* 100 α % point in the probability density function (pdf) of

$$t(c_1, \cdots, c_n) = \frac{\sum_1^n c_i Y_{(i)} - Y_0}{s}. \tag{2.1.1}$$

The class $\mathcal{C}$ of decision rules $D(c_1, \cdots, c_n)$ $(c_i \geqq 0, i = 1, \cdots, n, \sum_1^n c_i = 1)$ is defined as follows:

"Reject any observation x_0 from the given observations x_i, $i = 0, 1, \cdots, n$, if

$$\sum_{i=1}^{n} c_i x_{(i)} - x_0 > s t_\alpha(c_1, \cdots, c_n) \tag{2.1.2}$$

and accept otherwise, where $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are n ordered observations among $(x_1, x_2, \cdots, x_n)$. (The $n + 1$ observations x_i, $i = 0, 1, \cdots, n$, taken

from normal populations are again assumed to be distinct.) Proceed as above for each of $n + 1$ observations separately, so that each of $n + 1$ observations in due turn takes the place of x_0 and the remaining ordered observations play the part of $x_{(1)}, \cdots, x_{(n)}$." Thus in the above procedure we may start with the largest observation among x_i, $i = 0, 1, \cdots, n$, as x_0 and work downwards. If any particular observation is rejected, all other observations smaller than this observation are automatically rejected.

For the sake of convenience, we shall denote the decision rule $D(c_1, \cdots, c_n)$ when (i) $c_i = 1/n$, $i = 1, \cdots, n$, by $\bar{D}$ and when (ii) $c_r = 1$, and $c_j = 0$, $j \neq r$ by $D(r)$, $1 \leqq r \leqq n$. The corresponding auxiliary statistics $t(c_1, \cdots, c_n)$ will be denoted by $\bar{t}$ and $t(r)$.

It may also be noted here that $\sum_1^n c_i x_{(i)} - x_0$ $(c_i \geqq 0, i = 1, \cdots, n, \sum_1^n c_i = 1)$ can be written in the alternative form $\sum_1^n c_i(x_{(i)} - x_0)$.

3. Some properties of class $\mathfrak{C}$.

3.1. *An inequality related to location parameters.*

THEOREM 3.1.1. *Suppose that* $F((x_1 - \mu_1) / \sigma_1, \cdots, (x_n - \mu_n) / \sigma_n)$ *is the cumulative distribution function* (cdf) *of n random variables* X_i, $i = 1, \cdots, n$, *and* $T(u_1, \cdots, u_n)$ *is a real-valued function of* u_i, $i = 1, \cdots, n$, *such that*

$$T(u_1 + \alpha_1, \cdots, u_n + \alpha_n) \geqq T(u_1, \cdots, u_n), \tag{3.1.1}$$

where $(\alpha_1, \cdots, \alpha_n)$ *is a set of real numbers and* $-\infty < u_i < \infty$, $i = 1, \cdots, n$. *If for an arbitrary constant k,*

$$P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix} \mu_1, \cdots, \mu_n \\ \sigma_1, \cdots, \sigma_n \end{matrix}\right]$$

denotes the probability of $T(X_1, \cdots, X_n) > k$ *when* $X_1, \cdots, X_n$ *have the* cdf $F((x_1 - \mu_1) / \sigma_1, \cdots, (x_n - \mu_n) / \sigma_n)$, *then*

$$P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix} \mu_1 + \alpha_1, \cdots, \mu_n + \alpha_n \\ \sigma_1, \cdots, \sigma_n \end{matrix}\right] \geqq P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix} \mu_1, \cdots, \mu_n \\ \sigma_1, \cdots, \sigma_n \end{matrix}\right].$$

PROOF.

$$\begin{aligned} P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix} \mu_1 + \alpha_1, \cdots, \mu_n + \alpha_n \\ \sigma_1 \quad, \cdots, \quad \sigma_n \end{matrix}\right] &= P\left[T(X_1 + \alpha_1, \cdots, X_n + \alpha_n) > k \,\middle|\, \begin{matrix} \mu_1, \cdots, \mu_n \\ \sigma_1, \cdots, \sigma_n \end{matrix}\right] \\ &\geqq P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix} \mu_1, \cdots, \mu_n \\ \sigma_1, \cdots, \sigma_n \end{matrix}\right], \end{aligned}$$

since $T(X_1 + \alpha_1, \cdots, X_n + \alpha_n) \geqq T(X_1, \cdots, X_n)$ by hypothesis. Q.E.D.

From this theorem the following corollaries readily follow:

COROLLARY 3.1.1. *If* (3.1.1) *is satisfied for all*

$$\alpha_i \geqq 0, i = 1, \cdots, n, \quad then \quad P[T(X_1, \cdots, X_n) > k]$$

is a nondecreasing function of each μ_i, $i = 1, \cdots, n$.

COROLLARY 3.1.2. *If*

$$T(u_1 + \alpha_1, \cdots, u_n + \alpha_n) > T(u_1, \cdots, u_n),$$

when $\alpha_i \geqq 0$ *and* $\alpha_i > 0$ *for at least one* i, $1 \leqq i \leqq n$, *and if the* cdf *of*

$$T(X_1, \cdots, X_n)$$

assigns a positive measure to every nondegenerate interval, then

$$P[T(X_1, \cdots, X_n) > k],$$

where k *is an arbitrary constant, is an increasing function of each* μ_i, $i = 1, \cdots, n$.

COROLLARY 3.1.3. *Any strictly monotonic functional of the* cdf *of*

$$T(X_1, \cdots, X_n),$$

which satisfies the conditions of Corollary 3.1.2, *is an increasing function of* μ_i, $i = 1, \cdots, n$.

EXAMPLE 1. Consider the pdf

$$f\left(\frac{x_1 - \mu_1}{\sigma_1}, \cdots, \frac{x_n - \mu_n}{\sigma_n}\right) = \prod_{i=1}^{n} (\sigma_i \sqrt{2\pi})^{-1} e^{-(x_i - \mu_i)^2/2\sigma_i}$$

and $T(X_1, \cdots, X_n) = (\sum_1^n c_i X_{(i)})^{2r+1}$, where $r = 0, 1, 2, \cdots$ and $c_i \geqq 0$, $i = 1, \cdots, n$, and $c_i > 0$ for at least one i.

Here the conditions of Corollary 3.1.2 are easily verified and it follows that

$$P\left[\left(\sum_{i=1}^{n} c_i X_{(i)}\right)^{2r+1} > k\right], \qquad r = 0, 1, 2, \cdots,$$

is an increasing function of each μ_i, $i = 1, \cdots, n$. This result for the particular case $r = 0$ will be used in Section 3 in proving the properties of unbiasedness and gradation for the class $\mathcal{C}$ of decision rules as defined in Section 2.1.

It is well known that if $F(x)$ is the cdf of a random variable X, then expectation $E(X)$, if it exists, is a strictly monotonic functional of F (cf. [6], p. 152–153; [12], p. 189). Hence we get from Corollary 3.1.3 that

$$E\left(\sum_{i=1}^{n} c_i X_{(i)}\right)^{2r+1}, \qquad r = 0, 1, 2, \cdots,$$

is an increasing function of each of μ_i, $i = 1, \cdots, n$. Much more complicated functions can be constructed (cf. [13], pp. 25–26) having a similar property.

3.2. *Property of unbiasedness.* Let $\Omega(\mu_1, \cdots, \mu_n; \sigma)$ denote the set of normal populations $N(\mu_i, \sigma^2)$, $i = 1, 2, \cdots, n$. Suppose that $X_{(1)} < \cdots < X_{(n)}$ are n order statistics from $\Omega(\mu_1, \cdots, \mu_n; \sigma)$ when one random observation from each

of these n normal populations with *common* variance equal to σ^2 is taken. Let X_0 be another independent variate obeying $N(\mu_0, \sigma^2)$. According to our decision rule $D(c_1, \cdots, c_n)$ as defined in Section 2.1, the probability of rejecting x_0 will then be given by

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha (c_1, \cdots, c_n)\right]$$

$$= \int_0^\infty ds \int_{-\infty}^{\infty} dx_0 \int \cdots_A \int p(s) e^{-(x_0-\mu_0)^2/2\sigma^2} \tag{3.2.1}$$

$$\cdot g(x_{(1)}, \cdots, x_{(n)} \mid \mu_1, \cdots, \mu_n)\, dx_{(1)} \cdots dx_{(n)},$$

where

$$A = \begin{pmatrix} -\infty < x_{(1)} < \cdots < x_{(n)} < \infty \\ \sum_{i=1}^{n} c_i x_{(i)} > x_0 + st_\alpha(c_1, \cdots, c_n) \end{pmatrix}, \tag{3.2.2}$$

$g(x_{(1)}, \cdots, x_{(n)} \mid \mu_1, \cdots, \mu_n)$ represents the pdf of $X_{(1)}, \cdots, X_{(n)}$ from $\Omega(\mu_1, \cdots, \mu_n; \sigma)$ and

$$p(s) = \frac{\nu^{\nu/2}}{2^{(\nu-2)/2}\,\Gamma(\nu/2)\sigma^\nu} e^{-\nu s^2/2\sigma^2} s^{\nu-1}, \tag{3.2.3}$$

i.e., the pdf of sample standard deviation s based on $\nu = (k - 1)(n + 1)$ (cf. Section 1) degrees of freedom.

THEOREM 3.2.1. *$P[\sum_1^n c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)]$ is an increasing function of each $\mu_i' = \mu_i - \mu_0$, $i = 1, \cdots, n$.*

PROOF. Let $X_i' = X_i - \mu_0$, $i = 0, 1, 2, \cdots, n$. Since $\sum_1^n c_i = 1$,

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right] \tag{3.2.4}$$

$$= P\left[\sum_{i=1}^{n} c_i X'_{(i)} > X_0' + st_\alpha(c_1, \cdots, c_n)\right].$$

For fixed values of X_0' and s, the conditional value of

$$P\left[\sum_{i=1}^{n} c_i X'_{(i)} > X_0' + st_\alpha(c_1, \cdots, c_n)\right]$$

is an increasing function of $\mu_i' = \mu_i - \mu_0$, by Corollary 3.1.2 and Example 1 of Section 3.1. Since the distribution of X_0' and s does not involve the μ_i', it is now obvious that the (unconditional) value of

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right]$$

is an increasing function of each μ_i'. Q.E.D.

From this theorem an interesting property of $D(c_1, \cdots, c_n)$ follows.

Corollary 3.2.1. *The probability of rejecting any undesirable population (i.e., any population which has not the largest mean) is never less than the probability of rejecting the desirable population (i.e., that population having the largest mean).*

Proof. For the present proof let $\mu_{(0)} \geqq \cdots \geqq \mu_{(n)}$ denote the mean of the given $n + 1$ normal populations with common variance σ^2. By our decision rule $D(c_1, \cdots, c_n)$ the probability of rejecting the desirable population $N(\mu_{(0)}, \sigma^2)$ will depend on

$$P_{c_1,\cdots,c_n}(\mu_{(1)} - \mu_{(0)}, \cdots, \mu_{(n)} - \mu_{(0)}), \tag{3.2.5}$$

which is defined as the conditional probability of

$$\sum_{i=1}^{n} c_i Y_{(i)} > y_0 + st_\alpha(c_1, \cdots, c_n),$$

when y_0 and s are assumed to be held constant. Here $Y_{(i)}$, $i = 1, \cdots, n$, and Y_0 are defined as in Section 2.1. The probability of rejecting any undesirable population $N(\mu_{(i)}, \sigma^2)$, $i = 1, 2, \cdots, n$, will, on the other hand, involve

$$\begin{aligned} P_{c_1,\cdots,c_n}(\mu_{(0)} - \mu_{(i)}, \mu_{(1)} - \mu_{(i)}, \cdots, \mu_{(i-1)} - \mu_{(i)}, \mu_{(i+1)} - \mu_{(i)}, \cdots, \\ \mu_{(n)} - \mu_{(i)}). \end{aligned} \tag{3.2.6}$$

Comparing the arguments of $P_{c_1,\cdots,c_n}$ in (3.2.5) and (3.2.6) we notice that

$$\mu_{(j)} - \mu_{(i)} \geqq \mu_{(j)} - \mu_{(0)}, \qquad \mu_{(0)} - \mu_{(i)} \geqq \mu_{(i)} - \mu_{(0)},$$

where $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, n$; $j \neq i$. Thus we can make a one to one correspondence between the n arguments of $P_{c_1,\cdots,c_n}$ in (3.2.5) and (3.2.6) in such a way that no argument of $P_{c_1,\cdots,c_n}$ in (3.2.6) is less than the corresponding argument of $P_{c_1,\cdots,c_n}$ in (3.2.5). Hence from the monotonic behavior of $P_{c_1,\cdots,c_n}(\delta_1, \cdots, \delta_n)$ with regard to δ_i, $i = 1, \cdots, n$, it follows that the probability of rejecting any undesirable population $N(\mu_{(i)}, \sigma^2)$, $i = 1, \cdots, n$, is never less than the probability of rejecting the desirable population $N(\mu_{(0)}, \sigma^2)$. This property may be denoted by the *property of unbiasedness* which is therefore possessed by our decision rules $D(c_1, \cdots, c_n)$ $(c_i \geqq 0, \sum_1^n c_i = 1)$. It may also be noted that all the arguments of $P_{c_1,\cdots,c_n}$ in (3.2.5) are nonpositive and so (3.2.5) will not exceed $P_{c_1,\cdots,c_n}(0, \cdots, 0)$. This implies that the probability of rejecting the desirable population $N(\mu_{(0)}, \sigma^2)$ will not exceed the desired significance level α $(0 < \alpha < 1)$. Hence α will be the least upper bound of the probability of incorrect choice (i.e., not including the population with the largest mean in the selected group), whatever may be the population means. Thus any rule $D(c_1, \cdots, c_n)$ satisfies the fundamental requirement as stated in Section 1.

3.3. *Property of gradation.* From Theorem 3.2.1 it follows that

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right] \tag{3.3.1}$$

is a decreasing function of μ_0; when $\mu_0 \to -\infty$, the value of (3.3.1) is equal to 1 and when $\mu_0 \to +\infty$, the same value is equal to 0. It is easily seen from (3.2.1)

that (3.3.1) is a continuous function of μ_0. Hence corresponding to any assigned value γ $(0 < \gamma < 1)$ of (3.3.1) there exists a particular value $\mu_{0\gamma}$ of μ_0 for which (3.3.1) is exactly equal to γ. The value $\mu_{0\gamma}$ will clearly in general depend upon $\mu_1, \cdots, \mu_n, \sigma$ and $c_1, \cdots, c_n$ besides the assigned value γ, and if

$$\mu_1, \mu_2, \cdots, \mu_n$$

increase by a given constant Δ, then $\mu_{0\gamma}$ will also be increased by the same constant. In this situation we shall, therefore, find that

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right] \gtreqless \gamma, \tag{3.3.2}$$

according as $\mu_0 \lesseqgtr \mu_{0\gamma}$. This property will be designated as the *property of gradation.* We shall now study the nature of the unknown constant $\mu_{0\alpha}$ when γ is taken to be equal to α $(0 < \alpha < 1)$. It will be shown that $\mu_{0\alpha}$ for the decision rule $\bar{D}$ is very simple in form, but for other decision rules of class $\mathfrak{C}$ no such simple explicit expression for $\mu_{0\alpha}$ can be given. Let $\bar{X}_n = \sum_1^n X_{(i)} / n$. Then $\bar{X}_n$ will obey $N(\sum_1^n \mu_i / n, \sigma^2 / n)$. Let $Y_{(1)} < \cdots < Y_{(n)}$ be n order statistics derived from a random sample of size n from $N(0, \sigma^2)$. Also let $Y_0 = X_0 - \mu_0$, so that Y_0 obeys $N(0, \sigma^2)$. Clearly the distribution of $\bar{X}_n - \sum_1^n \mu_i / n$ is identical with that of $\bar{Y}_n = \sum_1^n Y_{(i)} / n$.

Under our decision rule $\bar{D}$ the probability of rejecting x_0 in a single rejection is equal to

$$P\,[\bar{X}_n - X_0 > s\bar{t}_\alpha] = P\left[(\bar{Y}_n - Y_0) + \left(\sum_{i=1}^{n} \mu_i / n - \mu_0\right) > s\bar{t}_\alpha\right]$$

$$\gtreqless P[\bar{Y}_n - Y_0 > s\bar{t}_\alpha] = \alpha,$$

according as $\mu_0 \lesseqgtr \sum_1^n \mu_i / n$. Thus for $\bar{D}$ we have the special property that the probability of rejecting any population whose mean is greater than the average of the remaining n population means is less than α and the probability of rejecting any population whose mean is not greater than the average of the remaining n population means is at least equal to α.

It is now shown that $\mu_{0\alpha}$ for the general decision rule $D(c_1, \cdots, c_n)$ is not in general equal to $E(\sum_1^n c_i X_{(i)})$, although we have just shown that this is true for $\bar{D}$.

The existence of $\mu_{0\alpha}$ (which is a function of $\mu_1, \cdots, \mu_n, \sigma; c_1, \cdots, c_n$ besides α) for which the property of gradation holds for the general decision rule has already been shown. This implies that

$$\alpha = P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right], \tag{3.3.3}$$

when $E(X_0) = \mu_{0\alpha}(\mu_1, \cdots, \mu_n, \sigma; c_1, \cdots, c_n, \alpha)$. It is shown that the assumption $\mu_{0\alpha} = E(\sum_1^n c_i X_{(i)})$ (which implies that $\mu_{0\alpha}$ is independent of α)

leads to a contradiction for the general case. The right-hand side of (3.3.3) can be written as

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - \mu_{0\alpha} - (X_0 - \mu_{0\alpha}) > st_\alpha(c_1, \cdots, c_n)\right]. \tag{3.3.4}$$

But $Y_0 = X_0 - \mu_{0\alpha}$ obeys $N(0, \sigma^2)$; hence by (3.3.3), (3.3.4), and the definition of $t_\alpha(c_1, \cdots, c_n)$ we get

$$\begin{aligned} P\left[\sum_{i=1}^{n} c_i Y_{(i)} - Y_0 > st_\alpha(c_1, \cdots, c_n)\right] &= \alpha \\ &= P\left[\sum_{i=1}^{n} c_i X_{(i)} - \mu_{0\alpha} - Y_0 > st_\alpha(c_1, \cdots, c_n)\right], \end{aligned} \tag{3.3.5}$$

where $Y_{(1)} < \cdots < Y_{(n)}$ are n order statistics obtained from a sample of size n from $N(0, \sigma^2)$. Hence it follows that, when $\mu_{0\alpha}$ is assumed to be independent of α, the distribution of $\sum_1^n c_i X_{(i)} - \mu_{0\alpha}$ and $\sum_1^n c_i Y_{(i)}$ must be identical. As a necessary condition for this we then have

$$\begin{aligned} E\left(\sum_{i=1}^{n} c_i Y_{(i)}\right) &= E\left(\sum_{i=1}^{n} c_i X_{(i)} - \mu_{0\alpha}\right) \\ &= E\left(\sum_{i=1}^{n} c_i X_{(i)}\right) - \mu_{0\alpha}. \end{aligned} \tag{3.3.6}$$

Hence if we assume that the unknown constant $\mu_{0\alpha}$ is $E(\sum_1^n c_i X_{(i)})$, then it will follow that

$$E\left(\sum_{i=1}^{n} c_i Y_{(i)}\right) = 0, \tag{3.3.7}$$

for an arbitrary set of c_i's such that $c_i \geq 0$ and $\sum_1^n c_i = 1$. The equation (3.3.7) does not, however, hold in general and hence we arrive at the conclusion that $\mu_{0\alpha}$ is not in general equal to $E(\sum_1^n c_i X_{(i)})$. We can, however, easily derive the value of $\mu_{0\alpha}$ for $D(c_1, \cdots, c_n)$ when $\mu_1 = \mu_2 = \cdots = \mu_n$. It can be easily shown (cf. [13], p. 70) that in such a situation $\mu_{0\alpha}$ must also be equal to μ_1.

4. Selection of an optimum rule.

4.1. In this section we shall assume that the number of degrees of freedom $(k - 1)(n + 1)$ of s (cf. Section 1) is so large that σ may be considered to be known. Under this restriction the rule $D(c_1, \cdots, c_n)$ as described in Section 2.1 requires the obvious modification that s should be replaced throughout by the population standard deviation σ.

It has been shown that the class $\mathfrak{C}$ of decision rules satisfies the fundamental requirement, i.e., the least upper bound of the probability of rejecting the population having the largest mean from the selected group is α $(0 < \alpha < 1)$, whatever may be the means of $n + 1$ given normal populations. If among the $n + 1$ population means all means except one are equal, then obviously it would be desirable to select that rule from the class $\mathfrak{C}$ which

(i) maximizes the probability of retaining in the selected group the population with the unequal mean if this is larger than the common mean of the other n populations; and

(ii) maximizes the probability of not retaining the population with the unequal mean if this is smaller than the mean of the other n populations.

In case (i) the population with the largest mean will be designated as the "best" population, and in case (ii) the population with the smallest mean will be called the "worst" population. Thus if $X_{(1)} < \cdots < X_{(n)}$ are assumed to have come from $N(0, \sigma^2)$ and X_0 from $N(\delta, \sigma^2)$, then our desirable rule should ensure largest probability (i) for retaining x_0 in this selected group if $0 < \delta < \infty$, or, (ii) for rejecting x_0 from the group if $-\infty < \delta < 0$. From what we have observed in Section 3.2 it is clear that the above rule will be optimum when $X_{(1)} < \cdots < X_{(n)}$ are assumed to arise from $N(\mu, \sigma^2)$ and X_0 from $N(\mu + \delta, \sigma^2)$, $-\infty < \mu < \infty$.

We shall now show that among the class $\mathcal{C}$ of decision rules the rule $\bar{D}$ maximizes (approximately) the probability of retaining the "best" population in the selected group. In an exactly analogous way it can be shown that $\bar{D}$ maximizes also the probability of rejecting the "worst" population from the group. To derive this result we shall first prove the following:

LEMMA 4.1.1. *Let $Y_{(1)} < \cdots < Y_{(n)}$ be n order statistics from $N(0, 1)$. Then $\sum_1^n Y_{(i)}/n = \sum_1^n Y_i/n$ has minimum variance among all $\sum_1^n c_i Y_{(i)}$ such that $\sum_1^n c_i = 1$.*

PROOF. We have

$$\operatorname{Var}\left(\sum_{i=1}^{n} c_i Y_{(i)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j v_{ij}, \tag{4.1.1}$$

where v_{ij} denotes the covariance between $Y_{(i)}$ and $Y_{(j)}$. Let the variance-covariance matrix of $Y_{(i)}$ and $Y_{(j)}$ $(i = 1, \cdots, n; j = 1, \cdots, n)$ be denoted by $\Sigma(n \times n)$.

To minimize (4.1.1) subject to the condition

$$\sum_{i=1}^{n} c_i = 1, \tag{4.1.2}$$

we get the following n equations

$$\sum_{j=1}^{n} c_j v_{ij} = \lambda, \qquad i = 1, \cdots, n, \tag{4.1.3}$$

where 2λ is used as Lagrangian multiplier. In matrix notation equations (4.1.3) can be written as

$$\Sigma \mathbf{c} = \lambda \mathbf{1}, \tag{4.1.3a}$$

where $\mathbf{c}'(1 \times n)$ and $\mathbf{1}'(1 \times n)$ denote the row vectors $(c_1, \cdots, c_n)$ and

$$(1, 1, \cdots, 1)$$

respectively. Since Σ is nonsingular, we get from (4.1.3a)

(4.1.4) $$\mathbf{c} = \lambda \Sigma^{-1} \mathbf{1}.$$

But it is known ([7], [8]) that

(4.1.5) $$\sum_{j=1}^{n} v_{ij} = 1.$$

Hence $\Sigma \mathbf{1} = \mathbf{1}$; this implies

(4.1.6) $$\Sigma^{-1} \mathbf{1} = \mathbf{1}.$$

By (4.1.2), (4.1.4), and (4.1.6) it follows that $c_i = 1/n$, $i = 1, \cdots, n$.

This completes the proof of the lemma.

The probability of retaining x_0 arising from the "best" population when $D(c_1, \cdots, c_n)$ is followed will clearly be given by

(4.1.7) $$\frac{n!}{(2\pi)^{(n+1)/2}\,\sigma^{n+1}} \int \cdots_{B} \int \exp\left[-(x_0 - \delta)^2 / 2\sigma^2 - \sum_{i=1}^{n} x_{(i)}^2 / 2\sigma^2 \right] \prod_{i=0}^{n} dx_{(i)},$$

where

$$B = \begin{bmatrix} -\infty < x_{(1)} < \cdots < x_{(n)} < \infty \\ \sum_{i=1}^{n} c_i x_{(i)} - x_0 < \sigma t_\alpha(c_1, \cdots, c_n) \\ -\infty < x_0 < \infty \end{bmatrix}.$$

Our object is to show that the expression (4.1.7) is (approximately) maximum for $\bar{D}$. The arguments given in [13], pp. 71–84 and [14] suggest that

$$u(c_1, \cdots, c_n) = \sum_{i=1}^{n} c_i Y_{(i)} - Y_0,$$

where the c_i's, $Y_{(i)}$'s and Y_0 have the same meaning as in Section 1, may be assumed to be normally distributed for all practical purposes whatever may be the value of n. Let the (approximate) normal distribution of $u(c_1, \cdots, c_n)$ be denoted by $N(\xi_c, \sigma_c^2)$. Henceforth we shall consider this distribution to be exactly normal and hence the result derived below is correct only approximately. For the special case when all c_i's are equal, i.e., $c_i = 1/n$, $i = 1, \cdots, n$, we shall write $\bar{u}$ for $u(1/n, \cdots, 1/n)$ and $N(\bar{\xi}, \bar{\sigma}^2)$ for the (exact) distribution of $\bar{u}$. By Lemma 4.1.1 we know that $\bar{\sigma}$ is the minimum among all σ_c, where $\sum_1^n c_i = 1$.

In the given situation $\sum_1^n c_i X_{(i)} - X_0 + \delta$ will have the (approximate) normal distribution $N(\xi_c, \sigma_c^2)$, where mean ξ_c and variance σ_c^2 are independent of δ. Hence

(4.1.8) $$v(c_1, \cdots, c_n) = \frac{\sum_1^n c_i X_{(i)} - X_0 + \delta - \xi_c}{\sigma_c}$$

will have standard normal distribution $N(0, 1)$.

Hence the expression in (4.1.7) can be written as

$$(4.1.9) \qquad (2\pi)^{-1/2} \int_{-\infty}^{[\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c + \delta]/\sigma_c} e^{-v^2/2}\, dv.$$

Also from the definition of $t_\alpha(c_1, \cdots, c_n)$ (cf. Section 1) it is now evident that

$$(4.1.10) \qquad (2\pi)^{-1/2} \int_{[\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c]/\sigma_c}^{\infty} e^{-v^2/2}\, dv = \alpha.$$

From (4.1.10) it follows that

$$(4.1.11) \qquad \frac{\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c}{\sigma_c} = \frac{\sigma \bar{t}_\alpha - \bar{\xi}}{\bar{\sigma}}.$$

From (4.1.9) it is easily seen that the probability of retaining x_0 in the selected group under the present situation is an increasing function of δ—a result which is a particular case of Corollary 3.1.2. Now for any arbitrary $\delta > 0$ the term in (4.1.9) will be maximum (when the c_i's are varied subject to the conditions $c_i \geqq 0$, $\sum_1^n c_i = 1$) when

$$(4.1.12) \qquad \frac{\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c + \delta}{\sigma_c} = \frac{\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c}{\sigma_c} + \frac{\delta}{\sigma_c}$$

is maximum. But $\sigma_c \geqq \bar{\sigma}$ (for all c_i's subject to the above restrictions) implies that $\delta/\bar{\sigma} \geqq \delta/\sigma_c$ for any $\delta > 0$. Hence by (4.1.11) and (4.1.12) it follows that (4.1.12) is maximum for $\bar{D}$. Thus the rule $\bar{D}$ may be taken as the optimum rule.

It is interesting to note the close similarity of this optimum rule $\bar{D}$ to the usual (Student's) t-statistic for which a desirable property has been recently derived by Bahadur [2] while studying two normal populations with a common variance.

5. Acknowledgements. The author wishes to express his deep gratitude to Professor R. C. Bose for his inspiring guidance in the course of this investigation and to Professors S. N. Roy and Wassily Hoeffding for their helpful suggestions and criticism.

REFERENCES

[1] Raghu Raj Bahadur, "On a problem in the theory of k populations", *Ann. Math. Stat.*, Vol. 21 (1950), pp. 362–375.

[2] R. R. Bahadur, "A property of the t-statistic", *Sankhyā*, Vol. 12 (1952), pp. 79–88.

[3] Robert E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–39.

[4] Robert E. Bechhofer, Charles W. Dunnett, and Milton Sobel, "A two-sample multiple decision procedure for ranking means of normal populations with unknown variances (Preliminary Report)" (abstract), *Ann. Math. Stat.*, Vol. 24 (1953), p. 136.

[5] Robert E. Bechhofer and Milton Sobel, "A sequential multiple decision procedure for ranking means of normal populations with known variances (Preliminary Report)" (abstract), *Ann. Math. Stat.*, Vol. 24 (1953), pp. 136–137.

[6] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press, Cambridge, (1934).

[7] HOWARD L. JONES, "Exact lower moments of order statistics in small samples from a normal distribution", *Ann. Math. Stat.*, Vol. 19 (1948), pp. 270–273.

[8] E. H. LLOYD, "Least squares estimation of location and scale parameters using order statistics", *Biometrika*, Vol. 39 (1952), pp. 88–95.

[9] EDWARD PAULSON, "A multiple decision procedure for certain problems in the analysis of variance", *Ann. Math. Stat.*, Vol. 20 (1949), pp. 95–98.

[10] EDWARD PAULSON, "On the comparison of several experimental categories with a control", *Ann. Math. Stat.*, Vol. 23 (1952), pp. 239–246.

[11] EDWARD PAULSON, "An optimum solution to the k-sample slippage problem for the normal distribution", *Ann. Math. Stat.*, Vol. 23 (1952), pp. 610–616.

[12] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation. 1. Validation of order statistics", *Ann. Math. Stat.*, Vol. 16 (1945), pp. 187–192.

[13] K. C. SEAL, "On a class of decision procedures for ranking means", Unpublished Thesis (1954), University of North Carolina, Chapel Hill.

[14] K. C. SEAL, "Approximate distribution of certain linear function of order statistics", *Sankhyā*, submitted for publication.

[15] L. H. C. TIPPETT, "The extreme individuals and the range of samples taken from a normal population", *Biometrika*, Vol. 17 (1925), pp. 151–164.

ORDERED FAMILIES OF DISTRIBUTIONS[1]

By E. L. Lehmann

University of California, Berkeley

1. Summary and introduction. A comparison is made of several definitions of ordered sets of distributions, some of which were introduced earlier by the author [7], [8] and by Rubin [10]. These definitions attempt to make precise the intuitive notion that large values of the parameter which labels the distributions go together with large values of the random variables themselves. Of the various definitions discussed the combination of two, (B) and (C) of Section 2, appears to be statistically most meaningful. In Section 3 it is shown that this ordering implies monotonicity for the power function of sequential probability ratio tests. In Section 4 the results are applied to obtaining tests that give a certain guaranteed power with a minimum number of observations. Finally, in Section 5, certain consequences are derived regarding the comparability of experiments in the sense of Blackwell [1].

2. Some definitions of order. Let $X = (X_1, \cdots, X_n)$ be a random vector with probability distribution P_θ, depending on a real parameter θ. In the problems occurring in applications such distributions are usually ordered in the sense, roughly speaking, that large values of θ lead on the whole to large values of the X's. This intuitive notion can be given a precise mathematical meaning in various ways, some of which we shall now describe.

(A′) For any $\theta < \theta'$ there exists a vector-valued function $f = (f_1, \cdots, f_n)$, depending in general on θ and θ', such that[2]

(i) $x \leqq f(x)$,

(ii) if X has distribution P_θ, then the distribution of $(f_1(X), \cdots, f_n(X))$ is $P_{\theta'}$.

This condition, which was used by the author in [7] and [8], states that one can pass from a random vector with distribution P_θ to one with distribution $P_{\theta'}$ by a transformation which increases all of the components of the vector. An example is the case of a location parameter θ where one can then put

$$f_i(x) = x_i + \theta' - \theta.$$

For technical reasons the following slightly weaker condition, which was given in [8], is sometimes more convenient.

(A) There exists a random vector Z and functions $g = (g_1, \cdots, g_n)$, $g' = (g_1', \cdots, g_n')$ such that

(i) $g(z) \leqq g'(z)$ for all z,

(ii) the distributions of $g(Z)$ and $g'(Z)$ are P_θ and $P_{\theta'}$ respectively.

Received September 14, 1954.

[1] This paper was prepared with the partial support of the Office of Naval Research.

[2] Here, as throughout, an inequality between two vectors means that this inequality holds for all the components.

A function ϕ defined on an n-dimensional euclidean space is said to be increasing if $x \leqq x'$ implies $\phi(x) \leqq \phi(x')$; a set S is said to be *increasing* if its characteristic set function is, that is, if $x \varepsilon S$, $x \leqq x'$ implies $x' \varepsilon S$.

Condition (A′) is the special case in which $Z = X$, g is the identity function, and $g' = f$. Condition (A) clearly implies:

(B) If $\theta < \theta'$, then for every increasing set[3] S

$$P_\theta(S) \leqq P_{\theta'}(S), \tag{2.1}$$

and also the seemingly stronger

(B′) If $\theta < \theta'$, then for every increasing function[3] $\phi(x_1, \cdots, x_n)$

$$E_\theta\phi(X) \leqq E_{\theta'}\phi(X). \tag{2.2}$$

Actually, (B) and (B′) are equivalent. To see this, assume without loss of generality that ϕ is non-negative, and consider the approximation of ϕ by a sequence of nondecreasing simple functions

$$\phi_n(x) = \begin{cases} \dfrac{i-1}{2^n} & \text{for} \quad x \varepsilon S_i^{(n)} \\ n \quad \text{for} & x \varepsilon S_N^{(n)} \end{cases}$$

where

$$S_i^{(n)} = \left\{x: \frac{i-1}{2^n} \leqq \phi(x) < \frac{i}{2^n}\right\}, \qquad i = 1, \cdots, n\cdot 2^n,$$

$$S_N^{(n)} = \{x:\phi(x) > n\} \qquad N = n\cdot 2^n + 1.$$

Then it is seen that $E_\theta \phi_n(X)$ can be written in the form

$$\sum_{i=1}^{N} a_i P_\theta(S_i^{(n)} + S_{i+1}^{(n)} + \cdots + S_N^{(n)})$$

where the a_i are $\geqq 0$. But each set $S_i^{(n)} + \cdots + S_N^{(n)}$ is increasing, and it follows from condition (B) that $E_\theta \phi_n(X) \leqq E_{\theta'}\phi_n(X)$ and hence $E_\theta \phi(X) \leqq E_{\theta'}\phi(X)$.

A somewhat different condition supposes that all of the distributions P_θ possess probability densities with respect to a common σ-finite measure μ.

(C) If $\theta < \theta'$, the probability ratio

$$\frac{p_{\theta'}(x)}{p_\theta(x)} \tag{2.3}$$

is increasing.[4]

---

[3] Throughout, we restrict consideration to sets and functions which are Borel measurable.

[4] Probability densities being defined only up to sets of measure zero, condition (C) and similar conditions to be considered later, for example in connection with Theorem 3, should be interpreted to mean that there exist versions of these densities satisfying the condition in question. Furthermore, the condition is not meant to carry any implication as regards the points x at which both densities vanish.

Slightly more generally it is enough to assume the existence of real-valued functions $t_1, \cdots, t_k$ such that

$$\frac{p_{\theta'}(x)}{p_\theta(x)} = \frac{f_{\theta'}(t_1(x), \cdots, t_k(x))}{f_\theta(t_1(x), \cdots, t_k(x))}.$$

Then $t_1(x), \cdots, t_k(x)$ are sufficient statistics, and without loss of generality $f_\theta(t_1, \cdots, t_k)$ may be taken to be the generalized probability density of $T = (t_1(X), \cdots, t_k(X))$. Condition (C) is therefore essentially a generalization of one investigated by H. Rubin [10] to the effect that the ratio (2.3) is a monotone function of a real-valued statistic. We note the obvious lemma:

Lemma 1. *If for each x the density $p_\theta(x)$ is a differentiable function of θ, then a necessary and sufficient condition for* (C) *to hold is that $\partial / \partial\theta(\log p_\theta(x))$ be nondecreasing.*

It was pointed out above that (A) implies (B). The following examples show that (A) and (B) are not equivalent, and that in general (C) is not directly comparable to (A) or (B).

The situation is summarized in Table I in which the sign + or − indicates that the condition in question does or does not hold.

TABLE I

(A)	(B)	(C)	
+	+	+	Evidently possible
+	+	−	Example 2.1
+	−	+	Impossible since (A) implies (B)
+	−	−	Impossible since (A) implies (B)
−	+	+	Example 2.2
−	+	−	Example 2.3
−	−	+	Example 2.4
−	−	−	Evidently possible

Example 2.1. Let X be a random variable having a Cauchy distribution, with density

$$p_\theta(x) = \frac{1}{\pi}\frac{1}{1 + (x - \theta)^2}.$$

Then if $\theta < \theta'$, the transformation $f(x) = x + (\theta' - \theta)$ shows that (A) holds, and hence also (B). On the other hand, the ratio $p_{\theta'}(x) / p_\theta(x) \to 1$ as $x \to \pm\infty$, and hence obviously is not monotone.

Example 2.2. Let $n = 2$, and let the probability be concentrated on the four squares $A, \cdots, D$ indicated in Fig. 1a. The conditional distribution over each of the four squares is assumed uniform under both θ and θ'. The probabilities of the squares are given in Table II.

It is easily checked that (B) holds. Also

$$r(x_1, x_2) = \frac{P_{\theta'}(x_1, x_2)}{P_\theta(x_1, x_2)}$$

is larger in A than in either B, C, or D, so that (C) is satisfied. On the other hand, if there existed vectors $g(Z)$ and $g'(Z)$ with distributions P_θ and $P_{\theta'}$, and such that $g(z) \leqq g'(z)$ for all z, then $g'(z) \varepsilon C$ would imply $g(z) \varepsilon C$, and hence $P_{\theta'}(C) \leqq P_\theta(C)$. Thus (A) does not hold.

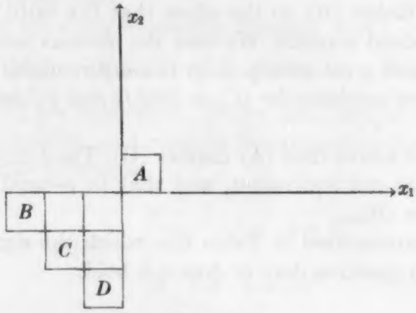

FIG. 1a

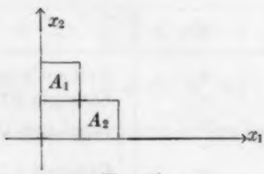

FIG. 1b

TABLE II

	P_θ	$P_{\theta'}$
A	3/16	12/16
B	6/16	1/16
C	1/16	2/16
D	6/16	1/16

Here the parameter θ takes on only two values. We obtain an example in which θ ranges over a continuum by means of the following lemma.

LEMMA 2. *Let P_0 and P_1 be two probability distributions, and let*

$$P_\theta = \theta P_1 + (1 - \theta)P_0, \qquad 0 \leqq \theta \leqq 1.$$

Then each of the conditions (A), (B), *and* (C) *holds for all* $0 \leqq \theta < \theta' \leqq 1$ *if and only if it holds for the pair* $\theta = 0$, $\theta' = 1$.

Proof. A direct calculation shows that if (B) or (C) holds for the pair $\theta = 0$, $\theta' = 1$, it holds for all $\theta < \theta'$. To prove this for (A), suppose that $f_i(Z)$ has distribution P_i $(i = 0, 1)$ and that $f_0(z) \leqq f_1(z)$ for all z. Consider a random variable U, uniformly distributed on [0, 1], and let

$$X_\theta = g(U, Z) = \begin{cases} f_0(Z) & \text{if} \quad U \leqq \theta \\ f_1(Z) & \text{if} \quad \theta < U \end{cases}$$

$$X_{\theta'} = g'(U, Z) = \begin{cases} f_0(Z) & \text{if} \quad U \leqq \theta' \\ f_1(Z) & \text{if} \quad \theta' < U. \end{cases}$$

Then X_θ and $X_{\theta'}$ have distributions P_θ and $P_{\theta'}$ respectively, and $g(u, z) \leqq g'(u, z)$ for all u and z.

The required example is now obtained by taking for P_0 and P_1 the probabilities denoted in the example by P_θ and $P_{\theta'}$, and by defining P_θ as in the lemma. This remark applies also to the examples that follow.

Example 2.3. In Fig. 1a of Example 2.2 replace the square A by two squares A_1, A_2 as indicated in Fig. 1b. Let the probabilities $P_\theta(A) = \frac{3}{16}$ and $P_{\theta'}(A) = \frac{12}{16}$ be divided among A_1 and A_2 so that

$$P_\theta(A_1) = \tfrac{2}{16}, \qquad P_\theta(A_2) = \tfrac{1}{16}; \qquad P_{\theta'}(A_1) = \tfrac{3}{16}, \qquad P_{\theta'}(A_2) = \tfrac{9}{16}.$$

Then as before (A) does not hold and (B) does. However, (C) now also does not hold since the ratio $r(x_1, x_2)$ has the value 2 in region C but only the value $\frac{3}{2}$ in region A_1.

Example 2.4. The (x_1, x_2)-plane is divided into 6 parts A_1, A_2, B_1, B_2, C_1, C_2 as indicated in Fig. 2. The probability ratio and the probabilities under θ and θ' of the six sets are given in Table III.

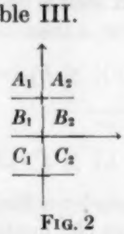

Fig. 2

TABLE III

	P_θ	$P_{\theta'}$	$r(x_1, x_2)$
A_1	.27	.45	5/3
A_2	.03	.05	
B_1	.30	.30	1
B_2	.10	.10	
C_1	.03	.01	1/3
C_2	.27	.09	

It is seen that (C) holds. On the other hand $P_\theta(A_2 + B_2 + C_2) = .40$, $P_{\theta'}(A_2 + B_2 + C_2) = .24$, so that (B) and hence (A) is not satisfied.

The rather chaotic state of things indicated by these examples is replaced by a much simpler one if the components of the vector X are independent, though not necessarily identically distributed.

THEOREM 1. *If $X_1, \cdots, X_n$ are independent, then*

$$(C) \rightarrow (B) \rightleftarrows (A).$$

PROOF. Consider first the case $n = 1$. Suppose that (B) holds, and let F_θ and $F_{\theta'}$ denote the cumulative distribution functions of the distributions P_θ and $P_{\theta'}$ respectively. If $g(z) = F_\theta^{-1}(z)$, $g'(z) = F_{\theta'}^{-1}(z)$, and Z is uniformly distributed on [0, 1], then $g(Z)$ and $g'(Z)$ have the distributions F_θ and $F_{\theta'}$ respectively. That $g(z) \leqq g'(z)$ follows from the fact that $F_{\theta'}(x) \leqq F_\theta(x)$ for all x since (B) is assumed to hold.

To show that (C) implies (B) when $n = 1$, let $r(x) = p_{\theta'}(x) / p_\theta(x)$. Given any constant k there exists a number ρ between 0 and 1 such that

$$P_\theta\{X > k\} = P_\theta\{r(X) > r(k)\} + \rho P_\theta\{r(X) = r(k)\}. \tag{2.4}$$

It is then easily seen that (2.4) holds, with the same ρ, also when θ is replaced by θ'. Consider now the problem of testing θ against θ', at the level of significance α, which is the value of the probability (2.4). Then the critical function, given by

$$\phi(x) = \begin{cases} 1 & \text{if} \quad r(x) > r(k) \\ \rho & \text{if} \quad r(x) = r(k) \end{cases}$$

has size α, and is the most powerful level α test for testing θ against θ'. It follows by comparison with the test $\phi(x) \equiv \alpha$ that

$$P_\theta\{r(X) > r(k)\} + \rho P_\theta\{r(X) = r(k)\} \leqq P_{\theta'}\{r(X) > r(k)\} + \rho P_{\theta'}\{r(X) = r(k)\}$$

and hence that for each k,

$$P_\theta\{X > k\} \leqq P_{\theta'}\{X > k\}.$$

The same relation for $X \geqq k$ follows by a limiting argument.

Suppose now that $n > 1$ and that (B) holds. Then in particular

$$P_\theta\{X_i > k\} \leqq P_{\theta'}\{X_i > k\} \qquad \text{for all } k \tag{2.5}$$

and it follows from the case $n = 1$ that (A) is satisfied.

Finally let $n > 1$, $p_\theta(x_1, \cdots, x_n) = f_\theta^{(1)}(x_1) \cdots f_\theta^{(n)}(x_n)$, and assume (C) to be satisfied. Then for each i, $f_{\theta'}^{(i)}(x_i) / f_\theta^{(i)}(x_i)$ is nondecreasing in x_i as is seen by holding the other coordinates fixed. It follows from the case $n = 1$ that (2.5) holds, and the proof is complete.

We shall in the present paper be mainly concerned with families of distributions that are ordered in the sense that both conditions (B) and (C) hold. It is a consequence of Theorem 1 that this is the case in particular if $X_1, \cdots, X_n$ is

a sample from a univariate distribution with density $f_\theta(x)$ where for $\theta < \theta'$ the ratio $f_{\theta'}(x) / f_\theta(x)$ is nondecreasing in x.

3. Monotonicity of the power function of some sequential tests. As a first application we consider the problem of testing sequentially the hypothesis $\theta \leqq \theta_0$ against the alternatives $\theta \geqq \theta_1$, where $\theta_0 < \theta_1$. Wald proposes as a solution the sequential probability ratio test, according to which observations are taken as long as

$$a < \sum_{i=1}^{m} \log \frac{p_{\theta_1}(x_i)}{p_{\theta_0}(x_i)} < b. \tag{3.1}$$

At the first violation of (3.1) the hypothesis is accepted or rejected according as the probability ratio is then $\leqq a$ or $\geqq b$.

Wald mentions ([12], p. 73) that in many important special cases the power function $\beta(\theta)$ of this test is an increasing function of θ. If a and b are adjusted so that $\beta(\theta_0) = \alpha$ and $\beta(\theta_1) = \beta$, this then implies that $\beta(\theta) \leqq \alpha$ for $\theta \leqq \theta_0$, and $\beta(\theta) \geqq \beta$ for $\theta \geqq \theta_1$, and hence satisfactory control of the probabilities of both kinds of error. The following result establishes such monotonicity for a large class of problems. The test treated is the *generalized probability ratio test*, where in (3.1) the constant boundaries a, b are replaced by variable boundaries, say a_m and b_m, and where some of the strong or weak inequality signs defining the test may be replaced by weak or strong ones respectively. This includes in particular the case of a single sample, or, more generally, of truncated sampling schemes if at some stage $a_m = b_m$.

THEOREM 2. *Let $X_1, X_2, \cdots$ be a sequence of random variables such that for all m the joint density $p_\theta^{(m)}(x_1, \cdots, x_m)$ of $X_1, \cdots, X_m$ satisfies* (B) *and* (C). *Then the power function $\beta(\theta)$ of any generalized probability ratio test is nondecreasing.*

PROOF. Let

$$z_m = \frac{p_{\theta_1}^{(m)}(x_1, \cdots, x_m)}{p_{\theta_0}^{(m)}(x_1, \cdots, x_m)}.$$

Then for $\theta < \theta'$ we have that for all k

$$P_\theta\{Z_m > k\} \leqq P_{\theta'}\{Z_m > k\}.$$

This follows from the fact that by (C), the set

$$\left\{(x_1, \cdots, x_m): \frac{p_{\theta_1}^{(m)}(x_1, \cdots, x_m)}{p_{\theta_0}^{(m)}(x_1, \cdots, x_m)} > k\right\}$$

is increasing, and that by (B) the probability of an increasing set is monotone in θ. Since Z_m is real-valued, there exists by Theorem 1 a real-valued function f_m such that $f_m(z) \geqq z$ for all z, and the distribution of $f_m(Z_m)$ is given by

$$P_\theta\{f_m(Z_m) \leqq u\} = P_{\theta'}\{Z_m \leqq u\} \qquad \text{for all } u.$$

Consider now the points $(1, Z_1)$, $(2, Z_2)$, $\cdots$ and the path they describe in the (i, Z)-plane. With the generalized probability ratio test, observations are taken

as long as this path lies within a certain prescribed band, and the hypothesis is accepted or rejected according as the path leaves the band for the first time through the upper or lower boundary. Now the path $\mathcal{C}'$ formed by the points $(1, f_1(Z_1))$, $(2, f_2(Z_2))$, $\cdots$ lies entirely above the path $\mathcal{C}$ formed by the points (i, Z_i), and hence whenever $\mathcal{C}$ leads to rejection by leaving the band through the upper boundary, so does $\mathcal{C}'$. But the probability of $\mathcal{C}$ and $\mathcal{C}'$ leading to rejection is exactly $\beta(\theta)$ and $\beta(\theta')$ respectively, which completes the proof.

It may be worth noting that use was made of condition (C) only for the pair of values (θ_0, θ_1).

Some simple applications of this theorem are to cases in which $X_1, X_2, \cdots$ are independently, identically distributed random variables, with probability density $f_\theta(x)$ for which $f_{\theta'}(x) / f_\theta(x)$ is nondecreasing in x whenever $\theta < \theta'$. In all such cases it follows from Theorems 1 and 2 that the power function of a generalized probability ratio test is nondecreasing.

EXAMPLE 3.1. Let the density of the X's be given by

$$f_\theta(x) = \theta g(x) + (1 - \theta)h(x), \qquad 0 \leqq \theta \leqq 1.$$

This is the situation in which the population under investigation is a mixture of two populations. In an experiment, for example, there may be the possibility of "gross errors" in addition to normal errors. Or it may be the problem of detecting the frequency of mutation of some gene, the effect of which is not directly observable. Since

$$\frac{f_{\theta'}(x)}{f_\theta(x)} = \frac{\theta'\left[\dfrac{g(x)}{h(x)} - 1\right] + 1}{\theta\left[\dfrac{g(x)}{h(x)} - 1\right] + 1},$$

it is seen that for $\theta < \theta'$ this ratio is increasing in x provided this is the case for $g(x) / h(x)$.

EXAMPLE 3.2. Let

$$f_\theta(x) = g(x - \theta). \tag{3.2}$$

Then (A) clearly holds without any restriction on the function g. On the other hand, (C) is exactly the condition of twice positivity of Schoenberg [11], a real-valued measurable function g being m times positive if, for every $k(=1, \cdots, m)$, $u_1 < u_2 < \cdots < u_k$, $v_1 < v_2 < \cdots < v_k$ implies that the determinant

$$\det \| g(u_i - v_j) \| \geqq 0.$$

A trivial specialization of Lemma 1 of [11] shows that a probability density g is twice positive if and only if (i) its domain of positivity is an interval (a, b), $-\infty \leqq a < b \leqq \infty$, (ii) the function $-\log g$ is convex (and hence automatically

continuous) in the open interval (a, b), and if g is correctly defined at the end points, as can always be achieved.[5]

As specific examples, let

$$g_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{(Normal)}$$

$$g_2(x) = e^{-x-e^{-x}}$$

$$g_3(x) = e^{-x} \quad \text{for} \quad x \geqq 0 \qquad \text{(Exponential)}$$

$$g_4(x) = 1 \quad \text{for} \quad -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2} \qquad \text{(Rectangular)}$$

$$g_5(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \qquad \text{(Logistic)}$$

$$g_6(x) = \tfrac{1}{2}e^{-|x|} \qquad \text{(Laplace)}$$

$$g_7(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \qquad \text{(Cauchy)}$$

In the first six of these cases $-\log g$ is convex while in the last it is not. A general class of densities of form (3.2) that satisfy condition (C) is formed by the cases in which g is a Polya frequency function. This class was defined and investigated by Schoenberg (see for example [11]) who showed these functions to be totally positive (that is, k times positive for all $k = 1, 2, \cdots$) and hence in particular twice positive.

EXAMPLE 3.3. Let

$$f_\theta(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right), \tag{3.3}$$

where g is an even function, and where, without loss of generality, one may restrict x to be nonnegative since the absolute values $| X_1 |, | X_2 |, \cdots$ form a set of sufficient statistics for θ. It is then seen as in the previous example, or can be deduced from it by transforming to $Y = \log X$, that (C) holds if and only if the domain of positivity of g is an interval (a, b) and $-\log g(e^x)$ is convex for $\log a < x < \log b$. This holds in the cases g_1, g_4, g_6 and g_7 of the previous example. Since the convexity of $-\log g(x)$ implies that of $-\log g(e^x)$ but not conversely, condition (C) in the case of an even function is more restrictive for a location parameter than for a scale parameter.

EXAMPLE 3.4. A well-known example, which satisfies also the stronger conditions investigated by Rubin [10], is that of an exponential family, with

$$f_\theta(x) = a(\theta)e^{xb(\theta)}h(x)$$

[5] The same condition was encountered in a slightly different context by Ruist, "Comparison for tests of nonparametric hypotheses," *Arkiv. för Mathematik*, Vol. 3 (1954), pp. 133–163. Logarithmically convex functions have also been considered by Artin in his "Einführung in die Theorie der Gammafunktion," *Hamburger Math. Einzelschriften*, No. 11, B. G. Teubner, Leipzig, (1931).

where $b(\theta)$ is a strictly increasing function of θ. This includes among others the binomial and Poisson families of distributions. It also includes the cases $f_\theta(y) = g(y - \theta)$ where g is one of the densities g_1, g_2 of Example 3.2, in the first case with $y = x$ and in the second with $y = e^{-e^{-x}}$. Still further special cases are obtained by putting $f_\theta(y) = (1/\theta)g(y/\theta)$ with g one of the functions g_1, g_3 or g_6 of Example 3.2 and $y = -x^2$, $y = -x$ and $y = -|x|$ respectively.

Without going into details we mention as further application of Theorem 2 some sequential tests of composite hypothesis, discussed among others by Wald [12], Cox [3], Johnson [4], such as the sequential t-tests or sequential analysis of variance tests. In those cases the variables $X_1, X_2, \cdots$ are dependent. That (C) holds follows from the fact that the noncentral t- and F- distributions satisfy (C) (see Section 4, Examples 4.3 and 4.4), while (B) is easily checked in all these cases.

4. Tests with guaranteed power. As another application consider the problem of testing that $\theta \leqq \theta_0$ against the alternatives $\theta \geqq \theta_1$ on the basis of $X = (X_1, \cdots, X_n)$. It is desired to find that test which, subject to

$$\beta(\theta) \leqq \alpha \qquad \text{for } \theta \leqq \theta_0, \tag{4.1}$$

maximizes the minimum power over $\theta \geqq \theta_1$, that is, which gives the greatest possible guaranteed power in that range. The solution to this problem is to determine a least favorable pair of distributions λ_0, λ_1 over the sets $\omega_0 = \{\theta : \theta \leqq \theta_0\}$ and $\omega_1 = \{\theta : \theta \geqq \theta_1\}$, and to reject the hypothesis when

$$\frac{\int_{\omega_1} p_\theta(x_1, \cdots, x_n)\, d\lambda_1(\theta)}{\int_{\omega_0} p_\theta(x_1, \cdots, x_n)\, d\lambda_0(\theta)} \geqq k. \tag{4.2}$$

If the family of distributions is ordered, it seems reasonable to expect that the least favorable distributions are those assigning probability 1 to the points θ_0 and θ_1 respectively, in which case (4.1) reduces to the probability ratio test

$$\frac{p_{\theta_1}(x_1, \cdots, x_n)}{p_{\theta_0}(x_1, \cdots, x_n)} > k. \tag{4.3}$$

It follows from Theorem 8.3 of [9] that (4.3) is the solution to the stated problem provided $\beta(\theta) \leqq \beta(\theta_0) = \alpha$ for $\theta \leqq \theta_0$ and $\beta(\theta) \geqq \beta(\theta_1)$ for $\theta \geqq \theta_1$. But this is certainly the case if $\beta(\theta)$ is nondecreasing. A sufficient condition for this is that both (B) and (C) hold since then the critical region (4.3) is increasing and hence its probability is a nondecreasing function of θ. (Actually, this is a special case of Theorem 2.) That (B) alone is not enough is seen, for example, in the Cauchy case. If $X_1, \cdots, X_n$ are independently, identically distributed with density $\pi^{-1} / (1 + (x - \theta)^2)$, it is seen that the region in which

$$\prod_{i=1}^{n} \frac{1 + (x_i - \theta_1)^2}{1 + (x_i - \theta_0)^2} > k$$

is a bounded set in n-space. Its probability therefore tends to zero as $\theta \to \infty$.

A limiting case, as $\theta_1 \to \theta_0$, of the property that the minimum power over $\theta \geqq \theta_1$ be a maximum is that of locally maximum power. Here one seeks the test which maximizes the derivative $\beta'(\theta_0)$ of the power function at $\theta = \theta_0$. If for any critical region w, the integral

$$\beta(\theta) = \int_w p_\theta(x)\, d\mu(x)$$

can be differentiated under the integral sign with respect to θ, the problem becomes that of maximizing

$$\beta'(\theta_0) = \int \frac{\partial}{\partial\theta} \log p_\theta(x)\bigg|_{\theta=\theta_0} p_{\theta_0}(x)\, d\mu(x)$$

subject to (4.1). If we again tentatively replace (4.1) by the side condition $\beta(\theta_0) = \alpha$, the best critical region by the Neyman-Pearson fundamental lemma is given by

$$\frac{\partial}{\partial\theta} \log p_\theta(x)\bigg|_{\theta=\theta_0} \geqq k. \tag{4.4}$$

If (C) holds, it was seen earlier that the left-hand side of (4.4) is a nondecreasing function of the x's. Hence it follows from (B) that $\beta(\theta) \leqq \beta(\theta_0) = \alpha$ for $\theta \leqq \theta_0$ and therefore that (4.4) is the desired result.

Let $X_1, \cdots, X_n$ be independently and identically distributed with density $f_\theta(x)$, which is either a mixture of two densities in proportion $\theta : 1 - \theta$, or where θ is a location or scale parameter, and suppose that the conditions of Examples 3.1–3.3 respectively are satisfied. Then the test maximizing the minimum power over $\theta \geqq \theta_0$ is given by the rejection region

$$\frac{f_{\theta_1}(x_1) \cdots f_{\theta_1}(x_n)}{f_{\theta_0}(x_1) \cdots f_{\theta_0}(x_n)} \geqq k \tag{4.5}$$

and the test maximizing the power locally by

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_\theta(x_i)\bigg|_{\theta=\theta_0} \geqq k'. \tag{4.6}$$

A uniformly most powerful one-sided test does of course usually not exist. A notable exception is the well-known case of the exponential family of Example 3.4.

As an illustration consider the case that $f_\theta(x) = g(x - \theta)$ where g is one of the densities g_i $(i = 1, \cdots, 7)$ of Example 3.2, and that $\theta_0 = 0$. For $i = 1, 2$ these are exponential families, and the test given by (4.6) is uniformly most powerful against the alternatives $\theta \geqq 0$. The same conclusion holds also for $i = 3$ since in that case $Y = \min(X_1, \cdots, X_n)$ is a sufficient statistic with density $n \exp[-n(y - \theta)]$ for $y \geqq \theta$. The case $i = 4$ is interesting in that again a uniformly most powerful one-sided test exists, although the minimal sufficient statistic is (Y, Z), with $Y = \min_i X_i$, $Z = \max_i X_i$, and hence two-dimensional. The explanation is that the statistic Y by itself is sufficient for $\theta \geqq 0$ when

attention is restricted to the part of the sample space that is possible when $\theta = 0$. In the case of the logistic distribution the locally most powerful test can be written down by substituting in (4.6). It is not uniformly most powerful, but is unbiased since (C) holds. For $i = 6$, when the sample is drawn from a Laplace distribution with unknown location parameter θ, the power function of a test may not be differentiable. However, it turns out that a locally most powerful test, in the natural sense of the term, still exists and, perhaps somewhat surprisingly, is given by the sign test, as will be shown in the appendix. Finally in the case $i = 7$, that of a Cauchy distribution, (C) does not hold, and the locally most powerful test does not seem to have a simple structure even when $n = 1$.

We now turn to some applications in which the variables $X_1, \cdots, X_n$ are not independent. Dependence may for example be introduced through the elimination of nuisance parameters by the principle of invariance or because the observable variables involve some common unobservable components, and the joint density of the x's will be a mixture of densities of independent variables. We first give a sufficient condition for (C) to hold in that case.

THEOREM 3. *Let $x = (x_1, \cdots, x_n)$ and let $g_\theta(x, \xi)$ be a family of densities depending on two real parameters θ and ξ and jointly measurable in x and ξ. For each θ, let λ_θ be a measure for ξ such that for all x, the integral*

$$p_\theta(x) = \int g_\theta(x, \xi)\, d\lambda_\theta(\xi)$$

exists. Then a sufficient condition for the family of densities $p_\theta(x)$ to satisfy (C) *is that for $\theta < \theta'$ condition* (C) *holds* (i) *for $g_\theta(x, \xi)$ when ξ is fixed and θ is taken as the parameter,* (ii) *for $g_\theta(x, \xi)$ when θ is fixed and ξ is taken as the parameter,* (iii) *for $d\lambda_\theta(\xi)$.*

Here in assumption (iii) the densities $d\lambda_\theta(\xi)$ and $d\lambda_{\theta'}(\xi)$ may be computed with respect to any σ-finite measure ν that dominates both of the given measures, since only the ratio of the densities matters. In the proof that follows and later in the paper we shall therefore denote this ratio by $d\lambda_{\theta'}(\xi) / d\lambda_\theta(\xi)$. This should not be taken to imply that $\lambda_{\theta'}$ is absolutely continuous with respect to λ_θ, but should be interpreted as a shorthand notation for $(d\lambda_{\theta'} / d\nu) : (d\lambda_\theta / d\nu)$.

PROOF. We must show that $x \leqq x'$ implies

$$\frac{\int g_\theta(x', \xi)\, d\lambda_\theta(\xi)}{\int g_\theta(x, \xi)\, d\lambda_\theta(\xi)} \leqq \frac{\int g_{\theta'}(x', \xi)\, d\lambda_{\theta'}(\xi)}{\int g_{\theta'}(x, \xi)\, d\lambda_{\theta'}(\xi)}. \tag{4.7}$$

Let Λ and Λ' be the probability distributions given by

$$d\Lambda(\xi) = \frac{g_\theta(x, \xi)\, d\lambda_\theta(\xi)}{\int g_\theta(x, \zeta)\, d\lambda_\theta(\zeta)}, \qquad d\Lambda'(\xi) = \frac{g_{\theta'}(x, \xi)\, d\lambda_{\theta'}(\xi)}{\int g_{\theta'}(x, \zeta)\, d\lambda_{\theta'}(\zeta)}.$$

These are the *a posteriori* distributions of ξ given x, corresponding to θ and θ' respectively. Then (4.7) may be rewritten as

$$\int \frac{g_\theta(x', \xi)}{g_\theta(x, \xi)}\, d\Lambda(\xi) \leqq \int \frac{g_{\theta'}(x', \xi)}{g_{\theta'}(x, \xi)}\, d\Lambda'(\xi). \tag{4.8}$$

By assumption (i) it is enough to prove that

$$\int \frac{g_\theta(x', \xi)}{g_\theta(x, \xi)}\, [d\Lambda'(\xi) - d\Lambda(\xi)] \geqq 0. \tag{4.9}$$

By assumption (iii) the ξ-axis can be divided into two mutually exclusive and exhaustive intervals S_- and S_+ such that S_- lies to the left of S_+ and $d\Lambda'(\xi) \,/\, d\Lambda(\xi)$ is $\leqq 1$ in S_- and $\geqq 1$ in S_+. We then have that the left-hand side of (4.9) equals

$$a \int_{S_-} [d\Lambda'(\xi) - d\Lambda(\xi)] + b \int_{S_+} [d\Lambda'(\xi) - d\Lambda(\xi)] \tag{4.10}$$

where a and b are mean values of $g_\theta(x', \xi) \,/\, g_\theta(x, \xi)$ in S_- and S_+ respectively so that by assumption (ii), $a \leqq b$. Since Λ and Λ' are probability measures, (4.10) becomes

$$(b - a) \int_{S_+} [d\Lambda'(\xi) - d\Lambda(\xi)] = (b - a) \int_{S_+} \left[\frac{d\Lambda'(\xi)}{d\Lambda(\xi)} - 1\right] d\Lambda(\xi) \geqq 0,$$

and was to be proved.

COROLLARY. *Let ξ be vector-valued, $\xi = (\xi_1, \cdots, \xi_s)$ say, and let*

$$p_\theta(x) = \int g_\theta(x, \xi)\, d\Lambda_\theta(\xi).$$

Suppose that the measure Λ_θ is the product of s linear measures $\Lambda_\theta = \lambda_\theta^{(1)} \times \lambda_\theta^{(2)} \times \cdots \times \lambda_\theta^{(s)}$ each of which satisfies condition (iii) *of Theorem 3, and that $g_\theta(x, \xi)$ satisfies condition* (i) *of this theorem. Suppose that condition* (ii) *is replaced by*

(ii′) *for each $j = 1, \cdots, s - 1$, the ratio*

$$\frac{g_\theta(x_1', \cdots, x_n', \xi_1', \cdots, \xi_j', \xi_{j+1}, \cdots, \xi_s)}{g_\theta(x_1, \cdots, x_n, \xi_1, \cdots, \xi_j, \xi_{j+1}, \cdots, \xi_s)}$$

is nondecreasing in $\xi_{j+1}, \cdots, \xi_s$ provided $x_i \leqq x_i'$ $(i = 1, \cdots, n)$ and $\xi_i \leqq \xi_i'$ $(i = 1, \cdots, j)$.

PROOF. It is seen from Theorem 3 by induction over j that

$$g_\theta^{(j)}(x_1, \cdots, x_n, \xi_{j+1}, \cdots, \xi_s) = \int \cdots \int g_\theta(x_1, \cdots, x_n, \xi_1, \cdots, \xi_s)\, d\lambda_\theta^{(1)}(\xi_1) \cdots d\lambda_\theta^{(j)}(\xi_j)$$

satisfies conditions (i) and (ii′), and this yields the desired result.

As an application we consider:

EXAMPLE 4.1. Let $U_1, \cdots, U_s$ be a sample from an unobservable random variable with density $f_\theta(u)$. What we observe are

$$X_{ij} = U_i + V_{ij}$$

where the V's are independently normally distributed with mean zero. For the moment we shall assume the variance of the V's to be known, and hence without loss of generality to be equal to 1. A typical example is the usual simplest model II problem in which θ is a scale parameter. We shall assume that $f_\theta(u)$ is an even function of u and that for $\theta < \theta'$ the ratio $f_{\theta'}(u) / f_\theta(u)$ is an increasing function of $|u|$, and consider the problem of testing $\theta \leqq \theta_0$ against $\theta \geqq \theta_1$. The joint density of the X's is given by

$$p_\theta(x) = \prod_{i=1}^{s} \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^n} \exp\left[-\frac{1}{2}\sum_{j=1}^{n} (x_{ij} - u_i)^2\right] f_\theta(u_i)\, du_i .$$

Therefore the absolute values of the means $\bar{x}_1, \cdots, \bar{x}_s$ constitute a set of sufficient statistics for θ, and we may restrict attention to them. Putting $y_i = \sqrt{n}\bar{x}_i$ and $\xi = \sqrt{n}u$, we obtain the joint density of the Y's as

$$p_\theta(y_1, \cdots, y_s) = C \prod_{i=1}^{s} \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}(y_i - \xi_i)^2\right] f_\theta(\xi_i / \sqrt{n})\, d\xi_i .$$

We shall now prove condition (C) for the density of the $|Y|$'s. Since we are dealing with a sample it is enough to check this for the case $s = 1$. We have

$$p_\theta(y) = Ce^{-\frac{1}{2}y^2} \int_0^{\infty} (e^{-\xi y} + e^{\xi y})e^{-\frac{1}{2}\xi^2} f(\xi / \sqrt{n})\, d\xi. \tag{4.11}$$

Condition (iii) of Theorem 3 is satisfied by assumption, and we need only check (i) and (ii) with

$$g_\theta(y, \xi) = e^{-\frac{1}{2}y^2}(e^{-\xi y} + e^{\xi y})e^{-\frac{1}{2}\xi^2} \qquad \text{for} \quad \xi > 0.$$

Since this is independent of θ, assumption (i) clearly holds. Examining (ii) we have

$$\frac{g_\theta(y', \xi)}{g_\theta(y, \xi)} = e^{-\frac{1}{2}(y'^2 - y^2)} \frac{e^{-\xi y'} + e^{\xi y'}}{e^{-\xi y} + e^{\xi y}} .$$

Now if $|y| \leqq |y'|$, it is easily checked that $(e^{-\xi y'} + e^{\xi y'}) / (e^{-\xi y} + e^{\xi y})$ is an increasing function of $|\xi|$, and this completes the proof of (C). It follows that the test which rejects when

$$\frac{p_{\theta_1}(y_1) \cdots p_{\theta_1}(y_s)}{p_{\theta_0}(y_1) \cdots p_{\theta_0}(y_s)} > k$$

where $p_\theta(y)$ is given by (4.11) maximizes the minimum power for testing $\theta \leqq \theta_0$ against $\theta \geqq \theta_1$.

We next consider the following somewhat more realistic case.

EXAMPLE 4.2. Suppose that the assumptions of Example 4.1 hold but that the variance σ^2 of the X's is unknown. We assume further that the unknown parameter in the distribution of the U's is a scale parameter, say τ. The problem is to test $\tau/\sigma \leqq \theta_0$ against $\tau/\sigma \geqq \theta_1$. Putting $\theta = \tau/\sigma$ the joint probability density of the X_{ij} is

$$\prod_{i=1}^{s} \int \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left[-\frac{1}{2\sigma^2}\sum (x_{ij} - u_i)^2\right] \frac{1}{\theta\sigma} f\left(\frac{u_i}{\theta\sigma}\right) du_i.$$

Here the statistics $V = \sum\sum (X_{ij} - \bar{X}_i)^2$, $\bar{X}_1, \cdots, \bar{X}_s$ are jointly sufficient. Putting $Y_i = \sqrt{n}\bar{X}_i$, $\xi_i = \sqrt{n}u_i$, the joint density of V and the Y's is given by

$$\frac{C}{\sigma^N} v^{(N-s)/2-1} \exp\left[-\frac{v}{2\sigma^2}\right] \prod_{i=1}^{s} \int \exp\left[-\frac{1}{2\sigma^2}(y_i - \xi_i)^2\right] \frac{1}{\theta\sigma^2} f\left(\frac{\xi_i}{\sqrt{n}\,\theta\sigma^2}\right) d\xi_i.$$

Now the problem of testing $\theta \leqq \theta_0$ against $\theta \geqq \theta_1$ remains invariant under multiplication of the Y's by a common positive constant a and of V by a^2, and there exists a solution to the given problem which is invariant under these transformations. We may therefore restrict attention to the maximal invariant $(z_1, \cdots, z_s)$ where $z_i = y_i / \sqrt{v}$. The joint density of the Z's is given by

$$C \int_0^{\infty} v^p e^{-v/2} \left[\prod_{i=1}^{s} \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}(z_i\sqrt{v} - \xi_i)^2\right] \frac{1}{\theta} f\left(\frac{\xi_i}{\theta\sqrt{n}}\right) d\xi_i\right] dv$$

$$= \int_0^{\infty} \cdots \int_0^{\infty} \frac{1}{\theta^s} f\left(\frac{\xi_i}{\theta\sqrt{n}}\right) \left[C \exp\left(-\tfrac{1}{2}\sum \xi_i^2\right) \int_0^{\infty} v^p e^{-v/2} \cdot \prod_{i=1}^{s} \left(e^{-\xi_i z_i \sqrt{v}} + e^{\xi_i z_i \sqrt{v}}\right) dv\right] d\xi_1 \cdots d\xi_s.$$

Denoting the expression in brackets by $g(z, \xi)$ we shall now show that $g(z', \xi) / g(z, \xi)$ is increasing in ξ for $z \leqq z'$, the other two conditions of Theorem 3 being satisfied as before. To prove that $g(z, \xi)$ has the desired property we apply once more Theorem 3 with z, v, ξ playing the role of θ, ξ, x in this order. The weight function for u being $d\lambda(v) = Cv^p e^{-v/2}$ independent of z, condition (iii) is satisfied. Putting

$$h_z(v, \xi) = C \exp\left[-\tfrac{1}{2}\sum \xi_i^2\right] \prod_{i=1}^{s} \left(e^{-\xi_i z_i \sqrt{v}} + e^{\xi_i z_i \sqrt{v}}\right)$$

it is enough to show that $h_{z'}(v, \xi) / h_z(v, \xi)$ is increasing in v and ξ and $h_z(v, \xi') / h_z(v, \xi)$ in ξ where the ξ's are assumed to be nonnegative and where $|z_i| \leqq |z_i'|$ for $i = 1, \cdots, s$. Now

$$\frac{h_z(v, \xi')}{h_z(v, \xi)} = \prod_i \frac{e^{-\xi_i' z_i \sqrt{v}} + e^{\xi_i' z_i \sqrt{v}}}{e^{-\xi_i z_i \sqrt{v}} + e^{\xi_i z_i \sqrt{v}}}$$

and each factor is increasing in v since $|\xi_i z_i| \leqq |\xi_i' z_i|$. Similarly $h_{z'}(v, \xi) / h_z(v, \xi)$ is increasing in v and $|\xi_i|$. Finally, condition (ii′) of the Corollary to Theorem 3

is checked in the same manner, and it therefore follows from this corollary that (C) holds for the density of the Z's.

From this and the fact that the density of the Z's is even in each of the variables it is seen that the most powerful invariant test for testing θ_0 against θ_1 has a rejection region which is increasing in $|z_1|, \cdots, |z_s|$. That the probability of such a region is increasing in θ is a consequence of the fact that condition (A) holds, θ being a scale parameter for the Z's.

As two further illustrations we prove that the noncentral t and F densities have monotone likelihood ratios, that is, satisfy (C), so that the associated tests have the minimax property discussed at the beginning of this section. The first of these results was earlier given by Kruskal [6]; the second was obtained by Rushton (personal communication) and by Meyer (in "An application of the invariance principle to the Student hypothesis," Technical Report No. 24, Department of Statistics, Stanford University, unpublished). A result containing these two as special cases was obtained about simultaneously with the present paper by Karlin ("On distributions $(p(x|\omega)$ for which $p(x|\omega_1) \cdot p(x|\omega_2)$ is monotone," Technical Report No. 26, Department of Statistics, Stanford University, unpublished), who considered densities of the form

$$g_\theta(x) = c(\theta)\phi(x) \int e^{\pm r(t)x} e^{\theta t} \, d\psi(t).$$

Example 4.3. Let $p_\theta(t)$ denote the noncentral t density with noncentrality parameter θ, (including as a particular case the central density for $\theta = 0$), that is, the density of Student's t statistic when the sample on which it is based is drawn from a normal distribution $N(\eta, \sigma^2)$. Then

$$p_\theta(t) = Ct^{-n} \int_0^\infty \exp\left[-\frac{n}{2}(w - \theta)^2\right] w^{n-1} e^{-w^2/2t^2} \, dw$$

where n is the sample size and $\theta = \eta/\sigma$. That $p_{\theta'}(t) \,/\, p_\theta(t)$ is an increasing function of t for $t \leqq 0$ follows directly from Theorem 3. For $t \geqq 0$ it can be seen by noting that Theorem 3 remains valid if the ratios considered in (ii) and (iii) are nonincreasing instead of nondecreasing, with the ratio considered in (i) remaining nondecreasing in x.

Example 4.4. The noncentral F-density with r and s degrees of freedom and noncentrality parameter θ is given by

$$p_\theta(u) = \sum_{k=0}^{\infty} P_\theta(k) h_{r+k,s+k}(u), \qquad u \geqq 0$$

where $h_{r+k,s+k}$ is the central F-density with $r + k$ and $s + k$ degrees of freedom, and where

$$P_\theta(k) = \theta^k e^{-\theta} \,/\, k!$$

is the Poisson probability with parameter θ. It again follows immediately from Theorem 3 that for $\theta < \theta'$ the ratio $p_{\theta'}(u) \,/\, p_\theta(u)$ is increasing in u.

We shall now mention some problems in which the conditions of Theorem 3 do not appear to be satisfied. In these situations it would be of interest to obtain basic densities f under which the probability ratio test for testing θ_0 against θ_1 maximizes the minimum power against $\theta \geqq \theta_1$. In all of these problems this is easily shown to be the case when f is the normal density.

PROBLEM 1. Let $X_1, \cdots, X_n$ be a sample from $(1/\theta)f((x - \xi) / \theta)$. Then the distribution of the differences $X_j - X_i$ depends only on θ. It follows from the Hunt-Stein theorem that for testing $\theta \leqq \theta_0$ there exists a test depending only on these differences and which maximizes the minimum power over $\theta \geqq \theta_1$. The problem mentioned then arises for the joint density of these differences, which is easily written down and which is of course independent of ξ. An elaboration of this problem is the case of two samples from densities $(1/\sigma)f((x - \xi) / \sigma)$ and $(1/\tau)f((y - \tau) / \tau)$ respectively, where $\theta = \sigma/\tau$.

PROBLEM 2. Let $X_1, \cdots, X_n$ be a sample from $(1/\sigma)f(x/\sigma - \theta)$. Here it is the ratios that play the role of the differences in Problem 1. In the two-sample version of this problem the samples came from $(1/\sigma)f((x - \xi) / \sigma)$ and $(1/\sigma)f((y - \eta) / \sigma)$, and $\theta = (\eta - \xi) / \sigma$.

PROBLEM 3. Let $X_1, \cdots, X_n$ be a sample from $f(x - \theta)$, where f is even and consider the problem of testing $|\theta| \leqq \theta_0$ against $|\theta| \geqq \theta_1$. Here one would expect the test that maximizes the minimum power to be given by the rejection region

$$\frac{f(x_1 - \theta_1) \cdots f(x_n - \theta_1) + f(x_1 + \theta_1) \cdots f(x_n + \theta_1)}{f(x_1 - \theta_0) \cdots f(x_n - \theta_0) + f(x_1 + \theta_0) \cdots f(x_n + \theta_0)} \geqq C.$$

This will be the case provided the probability of this region is an increasing function of $|\theta|$. The problem is to find conditions on f which would insure this.

5. Comparability of experiments. When a family of distributions $\{P_\theta\}$ is ordered, it seems reasonable to expect

(D). The pair of distributions (θ_0', θ_1') is more[6] informative than the pair (θ_0, θ_1) in the sense of Blackwell [1] provided $\theta_0' \leqq \theta_0 < \theta_1 \leqq \theta_1'$.

Let ϕ_α and ϕ_α' be the most powerful level α tests for testing θ_0 against θ_1 and θ_0' against θ_1' respectively. Then Blackwell showed in [2] that (θ_0', θ_1') is more informative than (θ_0, θ_1) if and only if $\beta_\alpha(\theta_1) \leqq \beta_\alpha'(\theta_1')$ for all α, where β_α and β_α' denote the power functions of ϕ_α and ϕ_α'.

A somewhat stronger property than (D) which one might also expect to hold in an ordered family is:

(E). Let $\theta_0 < \theta_1$ and let λ_0, λ_1 be any distributions over the sets $\theta \leqq \theta_0$ and $\theta \geqq \theta_1$ respectively. Then the pair of distributions $(\int p_\theta(x)\, d\lambda_0(\theta), \int p_\theta(x)\, d\lambda_1(\theta))$ is more informative than the pair $(p_{\theta_0}(x), p_{\theta_1}(x))$.

Clearly (E) is actually stronger than (D). As a trivial example, let X be normally distributed with unit variance and mean ξ and let $\theta_0 = \theta_0'$ correspond to $\xi = 0$, θ_1 to $\xi = -1$ and θ_1' to $\xi = +1$. Then (θ_0, θ_1) and (θ_0', θ_1') are equally

[6] Throughout we shall understand with Blackwell "more informative" in the weak sense of "at least as informative."

informative and both strictly more informative than $(\theta_0, \frac{1}{2}\theta_1 + \frac{1}{2}\theta_1')$. It would be interesting to know whether more natural examples of this phenomenon exist such as, for example, a family of densities $g(x - \theta)$ which satisfies (D) for all $\theta_0' \leqq \theta_0 < \theta_1 \leqq \theta'$ but for which (E) does not hold.

A condition equivalent to (E) is:

(E′). For every pair $\theta_0 < \theta_1$ and every α the power function $\beta(\theta)$ of the probability ratio test for testing θ_0 against θ_1 satisfies

$$\begin{aligned} \beta(\theta) &\leqq \beta(\theta_0) \quad &\text{for } \theta \leqq \theta_0, \\ \beta(\theta) &\geqq \beta(\theta_1) \quad &\text{for } \theta \geqq \theta_1. \end{aligned} \tag{5.1}$$

To see this, note that (E) states that at every level α, the pair of *a priori* distributions assigning probability 1 to θ_0 and θ_1 respectively is least favorable for testing $\theta \leqq \theta_0$ against $\theta \geqq \theta_1$. It follows from Theorem 3.10 of [13] that (E) implies (E′). The converse is also a special case of a well-known simple decision-theoretic result, or alternatively can be seen from the proof of Theorem 4. Since (E′) is a consequence of (B) + (C), so is (E). On the other hand, the following example shows that (B) is not enough to insure even (D).

EXAMPLE 5.1. Let X be uniformly distributed over the union of the two intervals $(\theta - \frac{3}{4}, \theta - \frac{1}{4})$, $(\theta + \frac{1}{4}, \theta + \frac{3}{4})$. Then (B) holds since θ is a location parameter. On the other hand, the pair of distributions $(\theta = 0, \theta = \frac{1}{2})$ is clearly strictly more informative than the pair $(\theta = 0, \theta = 1)$.

We shall finally show that (B) + (C) permit an even stronger conclusion than (E).

THEOREM 4. *Let $p_\theta(x)$ be a family of probability densities satisfying* (B) *and* (C). *Let (λ_0, λ_1) and (λ_0', λ_1') be two pairs of probability distributions for the parameter θ such that the three ratios $d\lambda_0 / d\lambda_0'$, $d\lambda_1 / d\lambda_0$, $d\lambda_1' / d\lambda_1$ are all nondecreasing. Then the experiment*

$$\left(\int p_\theta(x)\, d\lambda_0'(\theta), \int p_\theta(x)\, d\lambda_1'(\theta) \right)$$

is more informative than the experiment

$$\left(\int p_\theta(x)\, d\lambda_0(\theta), \int p_\theta(x)\, d\lambda_1(\theta) \right).$$

It is convenient to prove first the following lemma.

LEMMA 3. *Let $x = (x_1, \cdots, x_n)$, and let $p_\theta(x)$ be a family of densities satisfying conditions* (B) *and* (C). *Let λ, λ' be two probability measures for θ such that $d\lambda'(\theta) / d\lambda(\theta)$ is nondecreasing in θ. Then*

(i) $$\frac{\int p_\theta(x)\, d\lambda'(\theta)}{\int p_\theta(x)\, d\lambda(\theta)} \text{ is nondecreasing in } x,$$

(ii) *if $\phi(x)$ is nondecreasing in x,*

$$\int E_\theta \phi(x)\, d\lambda(\theta) \leqq \int E_\theta \phi(x)\, d\lambda'(\theta). \tag{5.1}$$

PROOF. (i) follows from Theorem 3, since $p_\theta(x') / p_\theta(x)$ is nondecreasing in x, and $d\lambda'(\theta) / d\lambda(\theta)$ is nondecreasing in θ. To see (ii), let $\psi(\theta) = E_\theta \phi(X)$. Then by (B'), $\psi(\theta)$ is nondecreasing and it is easily seen that

$$\int \psi(\theta)[d\lambda'(\theta) - d\lambda(\theta)] \geqq 0.$$

PROOF OF THEOREM 4. Let ϕ_α and ϕ_α' be the most powerful level α tests for testing $\int p_\theta(x)\, d\lambda_0(\theta)$ against $\int p_\theta(x)\, d\lambda_1(\theta)$, and $\int p_\theta(x)\, d\lambda_0'(\theta)$ against $\int p_\theta(x)\, d\lambda_1'(\theta)$ respectively. Let

$$\beta(\alpha) = \int E_\theta \phi_\alpha(x)\, d\lambda_1(\theta), \qquad \beta'(\alpha) = \int E_\theta \phi_\alpha'(x)\, d\lambda_1'(\theta)$$

denote the power of these two tests for their respective alternatives. Then the desired result follows if for all α we have $\beta(\alpha) \leqq \beta'(\alpha)$. It is seen from part (i) of Lemma 3 that the rejection functions ϕ_α and ϕ_α' are nondecreasing. Therefore, by part (ii) of the lemma

$$\int E_\theta \phi_\alpha(x)\, d\lambda_0'(\theta) \leqq \int E_\theta \phi_\alpha(x)\, d\lambda_0(\theta) = \alpha$$

so that ϕ is a level α test also for the hypothesis $\int p_\theta(x)\, d\lambda_0'(\theta)$, and

$$\int E_\theta \phi_\alpha(x)\, d\lambda_1'(\theta) \leqq \beta'(\alpha)$$

since $\beta'(\alpha)$ is the power of the most powerful level α test. Also, by part (ii) of the lemma

$$\beta(\alpha) = \int E_\theta \phi(x)\, d\lambda_1(\theta) \leqq \int E_\theta \phi(x)\, d\lambda_1'(\theta),$$

and the result follows.

In conclusion I should like to thank a referee of this paper for many very helpful suggestions.

6. Appendix. A property of the sign test. It was recently shown by Hoeffding and Rosenblatt ("The efficiency of tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 52–63) that the sign test is asymptotically most efficient for detecting a small shift in the distribution with density $\frac{1}{2}e^{-|x|}$. We shall show below that the sign test is in fact locally most powerful for testing $H: \theta = 0$ against the alternatives $\theta > 0$ when

$$p_\theta(x_1, \cdots, x_n) = \frac{1}{2^n} e^{-\Sigma |x_i - \theta|} \tag{6.1}$$

for any fixed sample size n. In this we shall restrict ourselves to levels of significance α at which the sign test can be carried out without randomization, that is, to one of the levels

$$\alpha_m = \sum_{k=0}^{m} \binom{n}{k} \Big/ 2^n, \qquad m = 0, 1, \cdots, n-1. \tag{6.2}$$

Since the power function $\beta(\theta)$ of a test of this hypothesis may not be differentiable, we shall state the optimum property of the sign test more precisely as follows.

Let $\beta^*(\theta)$ be the power function of the sign test at one of the levels α_m and let $\beta_\phi(\theta)$ be the power function of any other test ϕ of H at the same level. Then there exists Δ such that

$$\beta_\phi(\theta) < \beta^*(\theta) \qquad \text{for} \quad 0 < \theta < \Delta. \tag{6.3}$$

To prove this, let us denote by R_k $(k = 0, \cdots, n)$ the subset of the sample space in which k is of the X's are positive and $n - k$ are negative. The proof follows easily from the following lemma.

LEMMA 4. *Let $0 \leqq k < l \leqq n$ and let S_k, S_l be subsets of R_k and R_l respectively for which*

$$P_0(S_k) = P_0(S_l). \tag{6.4}$$

Then there exists $\Delta_{k,l}$ such that

$$P_\theta(S_k) < P_\theta(S_l) \qquad \textit{for} \quad 0 < \theta < \Delta_{k,l}\,. \tag{6.5}$$

PROOF. We note that

$$\frac{e^{-|x-\theta|}}{e^{-|x|}} = \begin{cases} e^{-\theta} & \text{if} \quad x < 0 \\ e^{2x-\theta} & \text{if} \quad 0 < x < \theta \\ e^{\theta} & \text{if} \quad \theta < x \end{cases}$$

and that $e^{-\theta} < e^{2x-\theta} < e^{\theta}$ if $0 < x < \theta$. Let $S_{l,\theta}$ denote the subset of S_l for which the l positive x's are all $> \theta$. Then

$$P_\theta(S_l) \geqq e^{(2l-n)\theta} P_0(S_{l,\theta}) + e^{-n\theta} P_0(S_l - S_{l,\theta})$$

$$P_\theta(S_k) \leqq e^{(2k-n)\theta} P_0(S_k).$$

Putting $\eta(\theta) = P_0(S_l - S_{l,\theta})$ and denoting the common value of (6.4) by γ, we therefore have

$$P_0(S_l) - P_0(S_k) \geqq e^{(2l-n)\theta}[\gamma - \eta(\theta)] + e^{-n\theta}\eta(\theta) - \gamma e^{(2k-n)\theta}.$$

This will be positive provided

$$\gamma[e^{(2l-n)\theta} - e^{(2k-n)\theta}] > \eta(\theta)[e^{-n\theta} - e^{(2k-n)\theta}].$$

Up to terms of order θ, the left- and right-hand sides are respectively $2\gamma(l - k)\theta$ and $\eta(\theta)(2k - n)\theta$. Since $\eta(\theta) \to 0$ as $\theta \to 0$, it follows that the desired inequality holds when θ is sufficiently small.

The result expressed by (6.3) is now an obvious consequence when the alternative test ϕ is nonrandomized. For consider any rejection region that does not consist of the upper tail of a sign test. Then it can be converted into a sign test of the same size by a finite number of steps, each of which consists in replacing an S_k by an S_l with $k < l$ which satisfies (6.4).

Only minor modifications of the argument are required in case the alternative test ϕ is randomized. In particular, in the lemma, the sets S_k and S_l are replaced by critical functions ϕ_k and ϕ_l over R_k and R_l respectively, such that

$$E_0 \phi_k(X_1, \cdots, X_n) = E_0 \phi_l(X_1, \cdots, X_n),$$

the conclusion being that

$$E_\theta \phi_k(X_1, \cdots, X_n) < E_\theta \phi_l(X_1, \cdots, X_n) \quad \text{for} \quad 0 < \theta < \Delta_{k,l}.$$

It is interesting to note that the sign test, being similar for testing $H:\theta = 0$ when the density (6.1) involves an unknown scale parameter, is also locally most powerful for that problem.

It should be mentioned finally that the above proof may be modified to show that the two-sided sign test maximizes $\frac{1}{2}[\beta(\theta) + \beta(-\theta)]$ for sufficiently small θ. This test is therefore locally most powerful among all tests that are symmetric with respect to the origin.

REFERENCES

[1] D. Blackwell, "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Univ. California Press, Berkeley, 1951, pp. 93–102.

[2] D. Blackwell, "Equivalent comparisons of experiments," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 265–272.

[3] D. R. Cox, "Sequential tests for composite hypotheses," *Proc. Cambridge Philos. Soc.*, Vol. 48 (1952), pp. 290–299.

[4] N. L. Johnson, "Some notes on the application of sequential methods in the analysis of variance," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 614–623.

[5] J. L. Hodges, Jr. and E. L. Lehmann, "Testing the approximate validity of statistical hypotheses," *J. Roy. Stat. Soc.*, to be published.

[6] W. Kruskal, "The monotonicity of the ratio of two noncentral t density functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 162–165.

[7] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[8] E. L. Lehmann, "Testing multiparameter hypotheses," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 541–552.

[9] E. L. Lehmann, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 20 (1950), pp. 1–26.

[10] H. Rubin, "A complete class of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 22 (1951), p. 608 (Abstract).

[11] I. J. Schoenberg, "On Polya frequency functions. I. The totally positive functions and their Laplace transforms," *J. d'Analyse Mathematique*, Vol. 1 (1951), pp. 331–374.

[12] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1947.

[13] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

ON AN APPLICATION OF KRONECKER PRODUCT OF MATRICES TO STATISTICAL DESIGNS[1]

MANOHAR NARHAR VARTAK

University of Bombay

1. Summary. By a statistical design (or simply, a design) we mean an arrangement of a certain number of "treatments" in a certain number of "blocks" in such a way that some prescribed combinatorial conditions are fulfilled. With every design is associated a unique matrix called the incidence matrix of the design (definitions, etc., in subsequent sections). In many instances, e.g., [7], [8], [10], [12], [16], information regarding certain kinds of designs such as BIB, PBIB designs is obtained from properties of the matrix NN' or of its determinant $|NN'|$ where N is the incidence matrix of the design under consideration. On the other hand in a few cases, such as [4], [5], [11], [14], [15], the incidence matrix N itself has been used to investigate properties of designs. This paper gives a method of using incidence matrices of known designs to obtain new designs.

In Section 2 we have defined the Kronecker product of matrices. This definition and some properties of the Kronecker product of matrices are given in [1]. Section 3 is devoted to a general discussion of an application of the concept of the Kronecker product of matrices to define the Kronecker product of designs. This section also contains two theorems which illustrate the use of the method of obtaining Kronecker products of designs. Definitions of some well-known designs are given in Section 4, which also contains a number of results giving explicit forms of certain Kronecker products. Finally some illustrations of a few results of Section 4 are given in Section 5.

2. The Kronecker product of matrices. Let

$$A = (a_{ij}), \qquad B = (b_{kl}), \qquad I_u, \qquad O_{m\times n} \tag{2.1}$$

be respectively an $m \times n$ matrix, a $p \times q$ matrix, the identity matrix of order u, the null or zero matrix of order $m \times n$, all defined over the set of non-negative integers. The Kronecker product of matrices A and B is defined as follows.

DEFINITION 2.1. The Kronecker product $A \times B$ of matrices A and B of (2.1) is defined by

$$A \times B = \begin{bmatrix} a_{11}\,B & a_{12}\,B & \cdots & a_{1n}\,B \\ a_{21}\,B & a_{22}\,B & \cdots & a_{2n}\,B \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1}\,B & a_{m2}\,B & \cdots & a_{mn}\,B \end{bmatrix}, \tag{2.2}$$

where $a_{ij}B$ $(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$ is itself a $p \times q$ matrix.

Received March 11, 1954.

[1] This work was supported by a Research Training Scholarship from the Government of India.

We shall always use an "$\times$" in the product of matrices to denote the Kronecker product. The ordinary product of matrices A and B (when it exists) will be denoted by $A \cdot B$ or AB.

It is clear from Definition 2.1 that the Kronecker product always exists and that $A \times B$ is an $mp \times nq$ matrix defined over the set of non-negative integers. Also it is obvious that the Kronecker product of two matrices reduces to the ordinary product if and only if one of the matrices is a scalar.

The result contained in the following theorem will be used later in Section 3.

THEOREM 2.1. *For any matrices A and B as in* (2.1) *we must have*

$$A \times B = P \cdot (B \times A) \cdot Q \tag{2.3}$$

where the matrices P and Q are obtained from the identity matrices I_{mp} and I_{nq} respectively by permuting rows and columns.

It should be noted that P and Q are nonsingular matrices whose elements consist only of 0's and 1's, and that the matrices P and Q are the same for any A and B as defined in (2.1).

A proof of Theorem 2.1 can be constructed from that of a similar result proved by Murnaghan [1], who gives various other properties of the Kronecker product.

3. The Kronecker product of designs. Let D_p, $p = 1, 2$, be a design in which v_p treatments are arranged in b_p blocks. Let N_p, the incidence matrix of the design D_p, be defined by

$$N_p = (n_{i_p j_p}^{(p)}), \qquad i_p = 1, 2, \cdots, v_p, \qquad j_p = 1, 2, \cdots, b_p, \tag{3.1}$$

where $n_{i_p j_p}^{(p)}$ is the number of times the i_pth treatment of D_p occurs in the j_pth block of D_p. Clearly $n_{i_p j_p}^{(p)}$ is a non-negative integer so that N_p is defined over the set of non-negative integers. Since a design uniquely determines its incidence matrix and vice versa, we may denote both a design and its incidence matrix by the same symbol. Also the treatments and blocks of a design correspond respectively to the rows and columns of the incidence matrix of the design.

Let N_1 and N_2 be the designs defined by (3.1). Then

$$N_{12} = N_1 \times N_2 \tag{3.2}$$

uniquely determines a design and so does

$$N_{21} = N_2 \times N_1 . \tag{3.3}$$

Theorem 2.1 at once leads to the following theorem.

THEOREM 3.1. *If N_1 and N_2 are designs defined by* (3.1), *then the designs N_{12} and N_{21} defined respectively by* (3.2) *and* (3.3) *are structurally the same, i.e., one of them can be obtained from the other by simply renaming the treatments and renumbering the blocks.*

This theorem enables us to designate the designs N_{12} and N_{21} by a common symbol N, the incidence matrix of N being taken to be $N_1 \times N_2$ or $N_2 \times N_1$, whichever is convenient.

Since the incidence matrix of the design N obtained above is the Kronecker

product of the incidence matrices of the designs N_1 and N_2, we may say that the design N is the Kronecker product of the designs N_1 and N_2.

We now examine a few matrices and the corresponding designs.

3(a). Let N_1 be a row n-vector

$$N_1 = (1\ 1\ \cdots\ 1) \tag{3.4}$$

there being n 1's on the right-hand side. The design N_1 clearly consists of n blocks each of size one, each block being treated by the same single treatment.

3(b). Let N_2 be a column m-vector

$$N_2 = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}, \tag{3.5}$$

there being m 1's on the right-hand side. The design N_2 is a single replication of m treatments in one block of size m.

If N_0 be any design, then with N_1 as in (3.4), we have

$$N^{(1)} = N_1 \times N_0 = (N_0\ N_0 \cdots N_0),$$

there being n N_0's on the right-hand side. This means that the design $N^{(1)}$ is nothing but n replications of the design N_0 as a whole. Again if N_0 be any design, then with N_2 as in (3.5), we have

$$N^{(2)} = N_0 \times N_2,$$

where clearly $N^{(2)}$ defines a design which is obtained from N_0 by replacing each treatment of N_0 by a group of m treatments. Also the rows of $N^{(2)}$ consist only of m repetitions of each row of N_0.

These two results can be combined into the following theorem.

THEOREM 3.2. *If N_0 be any design and if N_1 and N_2 be as defined in* (3.4) *and* (3.5) *respectively, then the designs $N^{(1)} = N_1 \times N_0$ and $N^{(2)} = N_0 \times N_2$ are respectively*

(i) *n replications of the design N_0 as a whole, and*

(ii) *the design obtained from N_0 by replacing each of its treatments by a group of m treatments so that the rows of $N^{(2)}$ consist only of m repetitions of each row of N_0.*

The following corollaries to Theorem 3.2 are trivial.

COROLLARY 3.2.1. *A randomized block design, N_3, with m treatments and n blocks, each block being a complete replication, is the Kronecker product*

$$N_3 = N_1 \times N_2 \tag{3.6}$$

of the designs N_1 and N_2 defined in (3.4) *and* (3.5) *respectively.*

COROLLARY 3.2.2. *If N_0 be any design and N_3 be the randomized block design defined in* (3.6), *then $N_3 \times N_0$ defines a design which contains n replications of the design derived from N_0 by replacing each of its treatments by a group of m treatments.*

3(c). The design corresponding to I_u, the identity matrix of order u, contains u treatments and u blocks each of size one, and the ith block contains a single plot to which the ith treatment is applied, $i = 1, 2, \cdots, u$.

The following corollary to Theorem 3.2 is also trivial.

COROLLARY 3.2.3. *With N_2 as defined in* (3.5), *we have*

$$N = I_u \times N_2,$$

which defines a design N useful for confounding with blocks the effects of certain treatment combinations of a factorial design when u and m have suitable values.

It may be noted that if N_0 be any design, then the Kronecker product $I_u \times N_0$ is always a disconnected design, and therefore no further illustrations involving I_u will be given.

4. Special cases of Kronecker products of designs. We first define a few designs.

4(a). Design N_2 is already defined in (3.5).

4(b). A balanced incomplete block (BIB) design N_{BIB} with parameters v^*, b^*, r^*, k^*, λ^* is defined to be the one in which the v^* treatments are arranged in the b^* blocks of size k^* each, such that

(i) the treatments in any block are all distinct,
(ii) each treatment is replicated r^* times, and
(iii) every pair of treatments occurs together in λ^* blocks (cf. [12]).

4(c). A partially balanced incomplete block (PBIB) design $N_{PBIB}^{(s)}$ with s associate classes and with parameters

$$v, \quad b, \quad r, \quad k, \quad n_i, \quad \lambda_i, \quad p_{jk}^i, \quad i, j, k = 1, 2, \cdots, s,$$

is defined as follows.

(i) There are v treatments arranged in b blocks each of size k such that each treatment is replicated r times and the treatments in any block are all distinct.

(ii) There can be established a relation of association between any two treatments, satisfying the following conditions.

(α) Two treatments are either 1st, 2nd, $\cdots$, or sth associates.

(β) Each treatment has n_i ith associates, $i = 1, 2, \cdots, s$.

(γ) Given any two treatments which are ith associates, the number of treatments which are common to the jth associates of the first and the kth associates of the second is p_{jk}^i and is independent of the pair of treatments with which we start.

(iii) Two treatments which are ith associates occur together in exactly λ_i blocks, $i = 1, 2, \cdots, s$ (cf. [8]).

PBIB designs with two associate classes have been extensively investigated by Bose and Shimamoto [9].

When the parameters $\lambda_1, \lambda_2, \cdots, \lambda_s$ of a PBIB design are not all different, the s associate classes of the PBIB design may not be all distinct. The following lemma, which is a modification of a remark by Rao [3], gives a criterion to determine whether the PBIB design has s or fewer distinct associate classes when its parameters $\lambda_1, \lambda_2, \cdots, \lambda_s$ are not all different.

LEMMA 4.1. *Let a* PBIB *design* $N^{(s)}_{\text{PBIB}}$ *with* s *associate classes and with parameters*

$$v,\quad b,\quad r,\quad k,\quad n_i\,,\quad \lambda_i\,,\quad p^i_{jk}\,,\qquad i,j,k = 1, 2, \cdots, s,$$

be such that $\lambda_1, \lambda_2, \cdots, \lambda_s$ *are not all different so that at least two of them are equal. Without loss of generality we can assume that* $\lambda_1 = \lambda_2$. *In this case the number of associate classes of the design* N^s_{PBIB} *can be reduced from* s *to* $s - 1$ *by combining its first two associate classes if and only if*

$$\text{(4.1)}\qquad \begin{bmatrix} \sum_{u,w=1}^{2} p^1_{uw} & \sum_{u=1}^{2} p^1_{u3} & \cdots & \sum_{u=1}^{2} p^1_{us} \\ \sum_{w=1}^{2} p^1_{3w} & p^1_{33} & \cdots & p^1_{3s} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{w=1}^{2} p^1_{sw} & p^1_{s3} & \cdots & p^1_{ss} \end{bmatrix} = \begin{bmatrix} \sum_{u,w=1}^{2} p^2_{uw} & \sum_{u=1}^{2} p^2_{u3} & \cdots & \sum_{u=1}^{2} p^2_{us} \\ \sum_{w=1}^{2} p^2_{3w} & p^2_{33} & \cdots & p^2_{3s} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{w=1}^{2} p^2_{sw} & p^2_{s3} & \cdots & p^2_{ss} \end{bmatrix}$$

Further if (4.1) *holds, then the parameters of the reduced* PBIB *design with* $s - 1$ *associate classes are*

$$v' = v,\qquad b' = b,\quad r' = r,\qquad k' = k,$$
$$n_1' = n_1 + n_2\,,\qquad n_2' = n_3\,, \cdots, n_{s-1}' = n_s\,,$$
$$\lambda_1' = \lambda_1 = \lambda_2\,,\qquad \lambda_2' = \lambda_3\,, \cdots, \lambda_{s-1}' = \lambda_s\,,$$

$$(p'^1_{yz}) = \begin{bmatrix} \sum_{u,w=1}^{2} p^t_{uw} & \sum_{u=1}^{2} p^t_{u3} & \cdots & \sum_{u=1}^{2} p^t_{us} \\ \sum_{w=1}^{2} p^t_{3w} & p^t_{33} & \cdots & p^t_{3s} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{w=1}^{2} p^t_{sw} & p^t_{s3} & \cdots & p^t_{ss} \end{bmatrix},$$

$$\text{(4.2)}\qquad (p'^x_{yz}) = \begin{bmatrix} \sum_{u,w=1}^{2} p^{x+1}_{uw} & \sum_{u=1}^{2} p^{x+1}_{u3} & \cdots & \sum_{u=1}^{2} p^{x+1}_{us} \\ \sum_{w=1}^{2} p^{x+1}_{3w} & p^{x+1}_{33} & \cdots & p^{x+1}_{3s} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{w=1}^{2} p^{x+1}_{sw} & p^{x+1}_{s3} & \cdots & p^{x+1}_{ss} \end{bmatrix}$$

where $t = 1$ *or* 2; $x = 2, 3, \cdots, s - 1$; $y, z = 1, 2, \cdots, s - 1$.

It follows that repeated applications of Lemma 4.1 to any PBIB design will ultimately give a PBIB design whose associated classes are all distinct.

The following results give the Kronecker products of various pairs of designs chosen from 4(a), 4(b), and 4(c).

THEOREM 4.1. (a) *The Kronecker product* $N = N_2 \times N_{\text{PBIB}}^{(s)}$ *of the design* N_2 *of* (3.5) *and a* PBIB *design* $N_{\text{PBIB}}^{(s)}$ *with* s *associate classes and with parameters*

$$v, \quad b, \quad r, \quad k, \quad n_i, \quad \lambda_i, \quad p_{jk}^i, \qquad i, j, k = 1, 2, \cdots, s,$$

is a PBIB *design with at most* $s + 1$ *associate classes.*

(b) *The design* N *defined above has* $s + 1$ *distinct associate classes if the design* $N_{\text{PBIB}}^{(s)}$ *has* s *distinct associate classes and* $\lambda_i < r$ *for all* $i = 1, 2, \cdots, s$.

(c) *In any case the parameters of the design* N *can be expressed in terms of those of the designs* N_2 *and* $N_{\text{PBIB}}^{(s)}$ *by the equations:*

$$\begin{aligned}
&v' = mv, \quad b' = b, \quad r' = r, \quad k' = mk, \\
&n_i' = mn_i, \quad n_{s+1}' = m - 1, \quad \lambda_i' = \lambda_i, \quad \lambda_{s+1}' = r,
\end{aligned}$$

$$(p_{yz}'^{i}) = \left[\begin{array}{c|c} m(p_{jk}^i) & (m-1)(\delta_{ji}) \\ \hline (m-1)(\delta_{ij}) & 0 \end{array}\right]$$

$$(p_{yz}'^{s+1}) = \left[\begin{array}{c|c} m(n_j\delta_{jk}) & O_{s\times 1} \\ \hline O_{1\times s} & m-2 \end{array}\right] \tag{4.3}$$

where $i, j, k = 1, 2, \cdots, s$; $y, z = 1, 2, \cdots, s + 1$; $\delta_{\alpha\beta} = 0$ *if* $\alpha \neq \beta$ *and* $\delta_{\alpha\beta} = 1$ *if* $\alpha = \beta$ *for all* $\alpha, \beta = 1, 2, \cdots$.

(d) *Lemma* 4.1 *can be applied to the cases in which the conditions in* (b) *above are not fulfilled.*

The essence of Theorem 4.1 appears in a paper by Zelen [17].

COROLLARY 4.1.1. *The Kronecker product* $N = N_2 \times N_{\text{BIB}}$ *of the design* N_2 *of* (3.5) *and a* BIB *design* N_{BIB} *with parameters* $v^*, b^*, r^*, k^*, \lambda^*$ *is a singular* GD *(group divisible) design with parameters*

$$\begin{aligned}
&v' = mv^*, \quad b' = b^*, \quad r' = r^*, \quad k' = mk^*, \\
&m' = v^*, \quad n' = m, \quad \lambda_1' = r^*, \quad \lambda_2' = \lambda^*.
\end{aligned} \tag{4.4}$$

It should be noted that singular GD designs can be obtained only by using Corollary 4.1.1 as was shown by Bose and Connor [8].

THEOREM 4.2. (a) *The Kronecker product* $N = N_{\text{PBIB}}^{(s)} \times N_{\text{PBIB}}^{(t)}$ *of two* PBIB *designs* $N_{\text{PBIB}}^{(s)}$ *and* $N_{\text{PBIB}}^{(t)}$ *with* s *and* t *associate classes respectively and with respective sets of parameters*

$$v_1, b_1, r_1, k_1, n_{1i_1}, \lambda_{1i_1}, p_{1j_1k_1}^{i_1}; \; i_1, j_1, k_1 = 1, 2, \cdots, s, \tag{4.5}$$

$$v_2, b_2, r_2, k_2, n_{2i_2}, \lambda_{2i_2}, p_{2j_2k_2}^{i_2}; \; i_2, j_2, k_2 = 1, 2, \cdots, t, \tag{4.6}$$

is a PBIB *design with at most* $t + s + ts$ *associate classes.*

(b) *The design N defined above has $t + s + ts$ distinct associate classes if*

(i) *the s associate classes of $N^{(s)}_{\text{PBIB}}$ and the t associate classes of $N^{(t)}_{\text{PBIB}}$ are all distinct,*

(ii) $\lambda_{1i_1} < r_1$, $\lambda_{2i_2} < r_2$, *and*

(iii) $r_1\lambda_{2i_2} \neq r_2\lambda_{1i_1}$ *for all* $i_1 = 1, 2, \cdots, s$ *and* $i_2 = 1, 2, \cdots, t$.

(c) *In any case the parameters of the design N can be expressed in terms of those of $N^{(s)}_{\text{PBIB}}$ and $N^{(t)}_{\text{PBIB}}$ by the equations:*

$$v' = v_1 \cdot v_2, \quad b' = b_1 \cdot b_2, \quad r' = r_1 \cdot r_2, \quad k' = k_1 \cdot k_2,$$

$$n'_{i_2} = n_{2i_2}, \quad n'_{t+i_1} = n_{1i_1}, \quad n'_{t+i_1+i_2s} = n_{2i_2} \cdot n_{1i_1},$$

$$\lambda'_{i_2} = r_1 \cdot \lambda_{2i_2}, \quad \lambda'_{t+i_1} = r_2 \cdot \lambda_{1i_1}, \quad \lambda'_{t+i_1+i_2s} = \lambda_{2i_2} \cdot \lambda_{1i_1},$$

$$(p'^{i_2}_{yz}) = \left[\begin{array}{c|c|c} (p^{i_2}_{2j_2k_2}) & O_{t\times s} & O_{t\times st} \\ \hline O_{s\times t} & O_{s\times s} & (\delta_{i_2j_2}) \times (n_{1j_1}\delta_{j_1k_1}) \\ \hline O_{st\times t} & (\delta_{j_2i_2}) \times (n_{1j_1}\delta_{j_1k_1}) & (p^{i_2}_{2j_2k_2}) \times (n_{1j_1}\delta_{j_1k_1}) \end{array}\right]$$

$$(4.7) \quad (p'^{t+i_1}_{yz}) = \left[\begin{array}{c|c|c} O_{t\times t} & O_{t\times s} & (n_{2j_2}\delta_{j_2k_2}) \times (\delta_{i_1j_1}) \\ \hline O_{s\times t} & (p^{i_1}_{1j_1k_1}) & O_{s\times st} \\ \hline (n_{2j_2}\delta_{j_2k_2}) \times (\delta_{j_1i_1}) & O_{st\times s} & (n_{2j_2}\delta_{j_2k_2}) \times (p^{i_1}_{1j_1k_1}) \end{array}\right]$$

$$(p'^{t+i_1+i_2s}_{yz}) = \left[\begin{array}{c|c|c} O_{t\times t} & (\delta_{j_2i_2}) \times (\delta_{i_1j_1}) & (p^{i_2}_{2j_2k_2}) \times (\delta_{i_1j_1}) \\ \hline (\delta_{i_2j_2}) \times (\delta_{j_1i_1}) & O_{s\times s} & (\delta_{i_2j_2}) \times (p^{i_1}_{1j_1k_1}) \\ \hline (p^{i_2}_{2j_2k_2}) \times (\delta_{j_1i_1}) & (\delta_{j_2i_2}) \times (p^{i_1}_{1j_1k_1}) & (p^{i_2}_{2j_2k_2}) \times (p^{i_1}_{1j_1k_1}) \end{array}\right]$$

where $i_1, j_1, k_1 = 1, 2, \cdots, s$; $i_2, j_2, k_2 = 1, 2, \cdots, t$;

$$y, z = 1, 2, \cdots, t + s + ts;$$

$\delta_{\alpha\beta} = 0$ *if* $\alpha \neq \beta$ *and* $\delta_{\alpha\beta} = 1$ *if* $\alpha = \beta$ *for all* $\alpha, \beta = 1, 2, \cdots$.

(d) *Lemma 4.1 can be applied to the cases in which the conditions in* (b) *above are not fulfilled.*

PROOF. Let us consider the Kronecker product of the design $N^{(s)}_{\text{PBIB}}$ and $N^{(t)}_{\text{PBIB}}$ in the form $N = N^{(s)}_{\text{PBIB}} \times N^{(t)}_{\text{PBIB}}$. Since $N^{(s)}_{\text{PBIB}}$ and $N^{(t)}_{\text{PBIB}}$ are PBIB designs with parameters (4.5) and (4.6) respectively, it follows that their incidence matrices $N^{(s)}_{\text{PBIB}}$ and $N^{(t)}_{\text{PBIB}}$ are of orders $v_1 \times b_1$ and $v_2 \times b_2$, respectively. Also the elements of these matrices consist only of 0's and 1's, there being r_1 1's in every row and

k_1 1's in every column of $N^{(s)}_{\text{PBIB}}$ and r_2 1's in every row and k_2 1's in every column of $N^{(t)}_{\text{PBIB}}$. In obtaining N from $N^{(s)}_{\text{PBIB}}$ and $N^{(t)}_{\text{PBIB}}$ we replace every 1 in $N^{(s)}_{\text{PBIB}}$ by the matrix $N^{(t)}_{\text{PBIB}}$ and every 0 in $N^{(s)}_{\text{PBIB}}$ by the null matrix $O_{v_2 \times b_2}$. From this it follows that the incidence matrix of N is a $v_1 \cdot v_2 \times b_1 \cdot b_2$ matrix whose elements consist only of 0's and 1's, there being $r_1 \cdot r_2$ 1's in every row and $k_1 \cdot k_2$ 1's in every column of N. This means that the parameters v', b', r', k' of N given by

$$(4.8) \qquad v' = v_1 \cdot v_2, \qquad b' = b_1 \cdot b_2, \qquad r' = r_1 \cdot r_2, \qquad k' = k_1 \cdot k_2$$

have their usual significance for the design N.

We shall now identify the various associate classes of a treatment of N. Let the first row of $N^{(s)}_{\text{PBIB}}$ correspond to the treatment Θ of $N^{(s)}_{\text{PBIB}}$, that of $N^{(t)}_{\text{PBIB}}$ to the treatment Θ' of $N^{(t)}_{\text{PBIB}}$, and that of N to the treatment θ of N. We shall identify the various associate classes of θ in N.

The row corresponding to Θ in $N^{(s)}_{\text{PBIB}}$ contains r_1 1's, all other elements in the row being 0. In obtaining N each of these r_1 1's is replaced by the matrix $N^{(t)}_{\text{PBIB}}$ and each of the 0's is replaced by the null matrix $O_{v_2 \times b_2}$. Hence the first v_2 rows of N contain exactly r_1 replications of the design $N^{(t)}_{\text{PBIB}}$ as a whole and nothing else. Consider one of these r_1 replications. Its first row, which corresponds to a section of θ in N, also corresponds to Θ' in $N^{(t)}_{\text{PBIB}}$. Consider the i_2th associates, $i_2 = 1, 2, \cdots, t$, of Θ' in $N^{(t)}_{\text{PBIB}}$ which occur in the replication of $N^{(t)}_{\text{PBIB}}$ under consideration. These are n_{2i_2} in number and each of them occurs together with Θ' in λ_{2i_2} blocks of the replication of $N^{(t)}_{\text{PBIB}}$. When we take into account all the r_1 replications of $N^{(t)}_{\text{PBIB}}$ in the first v_2 rows of N, we find that each of the n_{2i_2} i_2th associates of Θ' in $N^{(t)}_{\text{PBIB}}$, considered as treatments of N, will occur together with θ in $r_1 \cdot \lambda_{2i_2}$ blocks of N. We take these n_{2i_2} treatments of N to be the i_2th associates of θ in N. This identifies the first t associate classes of θ in N and we see that the parameters n'_{i_2}, λ'_{i_2} of N given by

$$(4.9) \qquad n'_{i_2} = n_{2i_2}, \qquad \lambda'_{i_2} = r_1 \lambda_{2i_2}, \qquad i_2 = 1, 2, \cdots, t,$$

have their usual significance for the design N.

Now consider the i_1th associates, $i_1 = 1, 2, \cdots, s$, of Θ in $N^{(s)}_{\text{PBIB}}$. These are n_{1i_1} in number, and each of them occurs together with Θ in λ_{1i_1} blocks of $N^{(s)}_{\text{PBIB}}$. Consider one of the n_{1i_1} i_1th associates of Θ. The row in $N^{(s)}_{\text{PBIB}}$ corresponding to this i_1th associate also contains r_1 1's and all other elements in the row are 0's. In obtaining N each of these r_1 1's is replaced by the matrix $N^{(t)}_{\text{PBIB}}$ and each of the 0's is replaced by the null matrix $O_{v_2 \times b_2}$. Hence the row in $N^{(s)}_{\text{PBIB}}$ corresponding to the i_1th associate of Θ under consideration gives rise to only r_1 replications in N of the design $N^{(t)}_{\text{PBIB}}$ as a whole and nothing else. Out of these r_1 replications of $N^{(t)}_{\text{PBIB}}$ only λ_{1i_1} can be paired off with similar replications of $N^{(t)}_{\text{PBIB}}$ in N arising out of the row corresponding to Θ, because Θ occurs together with any of its i_1th associates in λ_{1i_1} blocks of $N^{(s)}_{\text{PBIB}}$. Consider one of these λ_{1i_1} pairs of replications of $N^{(t)}_{\text{PBIB}}$. The first row in each component replication of $N^{(t)}_{\text{PBIB}}$ in this pair is identical with that corresponding to Θ'. One of these first rows is a section of that corresponding to θ in N; we define the treatment in N

corresponding to the other first row to be a $(t + i_1)$th associate of θ in N. Since Θ' is replicated r_2 times in $N^{(t)}_{\text{PBIB}}$, it follows that in the pair of replications of $N^{(t)}_{\text{PBIB}}$ considered above θ and the other treatment, which is defined to be a $(t + i_1)$th associate of θ in N, occur together in r_2 blocks of N. Taking into account the λ_{1i_1} pairs of replications of $N^{(t)}_{\text{PBIB}}$ brought to notice earlier, it follows that θ and its $(t + i_1)$th associate in N occur together in $r_2 \cdot \lambda_{1i_1}$ blocks of N. Also remembering that the number of the i_1th associates of Θ in $N^{(s)}_{\text{PBIB}}$ is n_{1i_1} we see that the number of the $(t + i_1)$th associates of θ in N is also n_{1i_1}. This identifies s more associate classes of θ in N and we see that the parameters n'_{t+i_1}, λ'_{t+i_1} of N given by

$$(4.10) \qquad n'_{t+i_1} = n_{1i_1}, \qquad \lambda'_{t+i_1} = r_2 \cdot \lambda_{1i_1}, \qquad i_1 = 1, 2, \cdots, s,$$

have their usual significance for the design N.

Consider again for a moment the above pair of replications of $N^{(t)}_{\text{PBIB}}$. Consider in particular that component $N^{(t)}_{\text{PBIB}}$ in this pair which contains the $(t + i_1)$th associate of θ in N. The first row of this $N^{(t)}_{\text{PBIB}}$ corresponds to Θ'. Consider the i_2th associates of Θ' in this $N^{(t)}_{\text{PBIB}}$. These, considered as treatments of N, are defined to be $(t + i_1 + i_2 s)$th associates of θ in N. In the replication of $N^{(t)}_{\text{PBIB}}$ under consideration they are n_{2i_2} in number and each of them occurs together with θ in λ_{2i_2} blocks of N. Remembering that there are λ_{1i_1} such replications of $N^{(t)}_{\text{PBIB}}$ corresponding to each of the n_{1i_1} i_1th associates of Θ in $N^{(s)}_{\text{PBIB}}$ it follows that there are $n_{1i_1} \cdot n_{2i_2}$ $(t + i_1 + i_2 s)$th associates of θ in N and each of them occurs together with θ in $\lambda_{1i_1} \cdot \lambda_{2i_2}$ blocks of N. This identifies ts further associate classes of θ in N, and the parameters $n'_{t+i_1+i_2 s}$, $\lambda'_{t+i_1+i_2 s}$ of N given by

$$(4.11) \qquad n'_{t+i_1+i_2 s} = n_{1i_1} \cdot n_{2i_2}, \qquad \lambda'_{t+i_1+i_2 s} = \lambda_{1i_1} \cdot \lambda_{2i_2}$$

are seen to have their usual significance for the design N.

Now, since $\sum_{i_1=1}^{s} n_{1i_1} = v_1 - 1$ and $\sum_{i_2=1}^{t} n_{2i_2} = v_2 - 1$ (cf. [2]), the number of treatments of N accounted for in the above identification of the various associate classes of θ in N is

$$\sum_{i_2=1}^{t} n_{2i_2} + \sum_{i_1=1}^{s} n_{1i_1} + \sum_{i_1=1}^{s} \sum_{i_2=1}^{t} n_{1i_1} \cdot n_{2i_2} = \left\{1 + \sum_{i_1=1}^{s} n_{1i_1}\right\}\left\{1 + \sum_{i_2=1}^{t} n_{2i_2}\right\} - 1$$
$$= v_1 v_2 - 1,$$

which together with θ exhausts all the $v_1 \cdot v_2$ treatments of N.

It may also be observed that if $1 \leqq i_1 \leqq s$ and $1 \leqq i_2 \leqq t$, then any integer m such that $t + s < m \leqq t + s + ts$ can be uniquely written in the form $m = t + i_1 + i_2 \cdot s$, that is, if $m = t + i_1 + i_2 s$ and if $m = t + i'_1 + i'_2 \cdot s$, then we must have $i_1 = i'_1$, and $i_2 = i'_2$. This fact ensures the uniqueness of the enumeration of the associate classes of θ in N, as described above.

This proves that the design N has at most $t + s + ts$ associate classes.

We now calculate the parameters p'^{z}_{xy}, $x, y, z = 1, 2, \cdots, t + s + ts$, of the design N.

Consider the treatment θ of N which corresponds to the first row of N. Let ϕ be an i_2th associate of θ in N, $i_2 = 1, 2, \cdots, t$. Then clearly the row in N corresponding to ϕ is contained in the first v_2 rows of N and these v_2 rows of N contain among themselves exactly r_1 replications of $N_{\text{PBIB}}^{(t)}$ as a whole and nothing else. The first row in any one of these replications of $N_{\text{PBIB}}^{(t)}$, which is a section of that corresponding to θ in N, corresponds to the treatment Θ' of $N_{\text{PBIB}}^{(t)}$. This replication of $N_{\text{PBIB}}^{(t)}$ also contains a row which is a section of the row corresponding to ϕ in N, and this row of $N_{\text{PBIB}}^{(t)}$ corresponds to the treatment Φ' of $N_{\text{PBIB}}^{(t)}$. Clearly Θ' and Φ' are i_2th associates of each other in $N_{\text{PBIB}}^{(t)}$. Now there are $p_{2j_2k_2}^{i_2}$ treatments of $N_{\text{PBIB}}^{(t)}$ which are in common with the j_2th associates of Θ' and the k_2th associates of Φ' in $N_{\text{PBIB}}^{(t)}$, for $j_2, k_2 = 1, 2, \cdots, t$. It is clear that exactly these $p_{2j_2k_2}^{i_2}$ treatments in the replication of $N_{\text{PBIB}}^{(t)}$ under consideration, considered as treatments of N, are those which are in common with the j_2th associates of θ and the k_2th associates of ϕ in N. Hence

$$(4.12) \qquad p_{j_2k_2}^{\prime i_2} = p_{2j_2k_2}^{i_2}, \qquad i_2, j_2, k_2 = 1, 2, \cdots, t.$$

Also observe that the first v_2 rows of N contain all the first t associate classes of θ, and also of ϕ, and only these. Hence we must have

$$(4.13) \quad p_{j_2u}^{\prime i_2} = 0; \qquad i_2, j_2 = 1, 2, \cdots, t; u = t + 1, t + 2, \cdots, t + s + ts.$$

Next consider the $(t + j_1)$th associates of θ in N. They are the treatments of N corresponding to those rows in N which correspond to Θ' in the replications of $N_{\text{PBIB}}^{(t)}$ arising from each of the j_1th associates of Θ in $N_{\text{PBIB}}^{(s)}$. Similarly the $(t + k_1)$th associates of ϕ in N are the treatments of N which correspond to those rows in N which correspond to Φ' in the replications of $N_{\text{PBIB}}^{(t)}$ arising from each of the k_1th associates of Θ in $N_{\text{PBIB}}^{(s)}$. Since the treatments Θ' and Φ' are distinct we must have

$$(4.14) \qquad p_{t+j_1,t+k_1}^{\prime i_2} = 0, \qquad j_1, k_1 = 1, 2, \cdots, s.$$

Again the $(t + k_1 + k_2 \cdot s)$th associates of ϕ in N are the k_2th associates of Φ' in the replications of $N_{\text{PBIB}}^{(t)}$ arising out of the k_1th associates of Θ in $N_{\text{PBIB}}^{(s)}$. To calculate the value of $p_{t+j_1,t+k_1+k_2s}^{\prime i_2}$ we have to count the number of treatments of N which are in common with the n_{1j_1} $(t + j_1)$th associates of θ and the $n_{1k_1} \cdot n_{2k_2}$ $(t + k_1 + k_2s)$th associates of ϕ in N. It is clear, from the way in which these associate classes are defined, that

$$p_{t+j_1,t+k_1+k_2s}^{\prime i_2} = 0 \quad \text{if} \quad j_1 \neq k_1,$$

$$p_{t+j_1,t+j_1+k_2s}^{\prime i_2} = 0 \quad \text{if} \quad i_2 \neq k_2,$$

$$p_{t+j_1,t+j_1+i_2s}^{\prime i_2} = n_{1j_1}.$$

It may be easily seen that the above relations can be written in the form

$$p_{t+j_1,t+k_1+k_2s}^{\prime i_2} = \delta_{i_2k_2} \cdot n_{1j_1} \cdot \delta_{j_1k_1}, \qquad j_1, k_1 = 1, 2, \cdots, s; k_2 = 1, 2, \cdots, t,$$

and since the indices j_1, k_1 have to run over their entire ranges before the index

k_2 can change its value, we can write the above equation in the matrix form

(4.15) $$(p'^{i_2}_{t+j_1,t+k_1+k_2s}) = (\delta_{i_2k_2}) \times (n_{1j_1}\delta_{j_1k_1}).$$

Finally let us consider the $(t + j_1 + j_2 \cdot s)$th associates of θ in N and the

$$(t + k_1 + k_2 \cdot s)\text{th}$$

associates of ϕ in N. The number of treatments of N in common with these two associate classes is $p'^{i_2}_{t+j_1+j_2\cdot s,t+k_1+k_2\cdot s}$. From the definitions of these two associate classes we find that

$$p'^{i_2}_{t+j_1+j_2s,t+k_1+k_2s} = 0 \quad \text{if} \quad j_1 \neq k_1,$$

$$p'^{i_2}_{t+j_1+j_2s,t+j_1+k_2s} = n_{1j_1}p^{i_2}_{2j_2k_2}.$$

These relations are easily seen to be equivalent to writing

$$p'^{i_2}_{t+j_1+j_2s,t+k_1+k_2s} = p^{i_2}_{2j_2k_2} \cdot n_{1j_1}\delta_{j_1k_1}, \quad j_1, k_1 = 1, 2, \cdots, s; j_2, k_2 = 1, 2, \cdots, t,$$

and since here also the indices j_1, k_1 have to run over their entire ranges before the indices j_2, k_2 can change their values, it follows that we can write the above equations in the matrix form

(4.16) $$(p'^{i_2}_{t+j_1+j_2s,t+k_1+k_2s}) = (p^{i_2}_{2j_2k_2}) \times (n_{1j_1}\delta_{j_1k_1}).$$

Combining the calculation in (4.12) to (4.16) and remembering that $p'^{i_2}_{yz} = p'^{i_2}_{zy}$, $i_2 = 1, 2, \cdots, t$; $y, z = 1, 2, \cdots, t + s + ts$, we get the first of the three matrices in (4.7).

Similar calculations will give the other two matrices in (4.7).

Thus the argument so far together with the results in (4.8) to (4.16) prove the statements (a) and (c) of Theorem 4.2.

Also from the way in which we have defined the various associate classes in N, we find that if the s associate classes of $N^{(s)}_{\text{PBIB}}$ and the t associate classes of $N^{(t)}_{\text{PBIB}}$ are all distinct, then the first t associate classes of N are all distinct, the next s associate classes of N are all distinct, and the last ts associate classes of N are all distinct. Further suppose that $\lambda_{1i_1} = r_1$ for some i_1, $1 \leqq i_1 \leqq s$. Then from (4.9) and (4.11) we find that $\lambda'_{i_2} = \lambda'_{t+i_1+i_2s}$, i_1 fixed; $i_2 = 1, 2, \cdots, t$; hence it may be possible to combine some of the corresponding associate classes. Similarly if $\lambda_{2i_2} = r_2$ for some i_2, $1 \leqq i_2 \leqq t$, then from (4.10) and (4.11) we find that $\lambda'_{t+i_1} = \lambda'_{t+i_1+i_2s}$, i_2 fixed; $i_1 = 1, 2, \cdots, s$; hence it may be possible to combine some of the corresponding associate classes. But if $\lambda_{1i_1} < r_1$ and $\lambda_{2i_2} < r_2$ for all $i_1 = 1, 2, \cdots, s$ and $i_2 = 1, 2, \cdots, t$, no such situation can arise and then the first t associate classes and the next s associate classes are distinct from the last ts associate classes of N. Finally if $r_1 \cdot \lambda_{2i_2} \neq r_2\lambda_{1i_1}$ for all $i_1 = 1, 2, \cdots, s$ and $i_2 = 1, 2, \cdots, t$, then the first t associate classes are distinct from the next s associate classes of N because of (4.9) and (4.10). This means that if the conditions in the statement (b) of Theorem 4.2 are satisfied, then the $t + s + ts$ associate classes of N are all distinct. This proves the statement (b) of Theorem 4.2.

Lastly the statement (d) of Theorem 4.2 is simply a provision for the cases to which the statement (b) of Theorem 4.2 does not apply.

This completes the proof of Theorem 4.2.

Although the following corollary is an obvious special case of Theorem 4.2 we state it separately because we shall require it for further investigation.

COROLLARY 4.2.1. (a) *The Kronecker product* $N = N_{(1)\text{BIB}} \times N_{(2)\text{BIB}}$ *of the two* BIB *designs* $N_{(1)\text{BIB}}$ *and* $N_{(2)\text{BIB}}$ *defined by the respective sets of parameters*

$$v_1^*, b_1^*, r_1^*, k_1^*, \lambda_1^*, \tag{4.17}$$

$$v_2^*, b_2^*, r_2^*, k_2^*, \lambda_2^*, \tag{4.18}$$

is a PBIB *design* N *with at most three associate classes.*

(b) *The three associate classes of the design* N *defined above are all distinct if* $r_1^* \cdot \lambda_2^* \neq r_2^* \cdot \lambda_1^*$.

(c) *In any case the parameters of the design* N *can be expressed in terms of those of* $N_{(1)\text{BIB}}$ *and* $N_{(2)\text{BIB}}$ *by the equations*

$$\begin{gathered} v' = v_1^* \cdot v_2^*, \quad b' = b_1^* \cdot b_2^*, \quad r' = r_1^* \cdot r_2^*, \quad k' = k_1^* \cdot k_2^*, \\ n_1' = v_2^* - 1, \quad n_2' = v_1^* - 1, \quad n_3' = (v_1^* - 1)(v_2^* - 1), \\ \lambda_1' = r_1^* \cdot \lambda_2^*, \quad \lambda_2' = r_2^* \cdot \lambda_1^*, \quad \lambda_3' = \lambda_1^* \cdot \lambda_2^*, \end{gathered} \tag{4.19}$$

$$(p_{yz}'^{1}) = \begin{bmatrix} v_2^* - 2 & 0 & 0 \\ 0 & 0 & v_1^* - 1 \\ 0 & v_1^* - 1 & (v_1^* - 1)(v_2^* - 2) \end{bmatrix},$$

$$(p_{yz}'^{2}) = \begin{bmatrix} 0 & 0 & v_2^* - 1 \\ 0 & v_1^* - 2 & 0 \\ v_2^* - 1 & 0 & (v_1^* - 2)(v_2^* - 1) \end{bmatrix},$$

$$(p_{yz}'^{3}) = \begin{bmatrix} 0 & 1 & v_2^* - 2 \\ 1 & 0 & v_1^* - 2 \\ v_2^* - 2 & v_1^* - 2 & (v_1^* - 2)(v_2^* - 2) \end{bmatrix}$$

where $y, z = 1, 2, 3$.

(d) *Lemma* 4.1 *can be applied to the cases in which the condition in* (b) *above is not fulfilled.*

We shall now obtain the conditions under which the Kronecker product N of two BIB designs $N_{(1)\text{BIB}}$ and $N_{(2)\text{BIB}}$, defined in Corollary 4.2.1.(a), is a PBIB design with only two distinct associate classes.

Since $\lambda_1^* < r_1^*$ and $\lambda_2^* < r_2^*$, it is clear that the first necessary condition is that $r_1^* \lambda_2^* = r_2^* \lambda_1^*$. In this case applying Lemma 4.1 to the first two matrices in (4.19) we find that the second necessary condition is that $v_1^* = v_2^*$. It is clear from Lemma 4.1 that these conditions are also sufficient for the design (4.19) to have only two distinct associate classes.

If the conditions $v_1^* = v_2^*$ and $r_1^* \cdot \lambda_2^* = r_2^* \cdot \lambda_1^*$ are satisfied, then from the relations among the parameters of BIB designs (cf. [11]), we must have $k_1^* = k_2^*$. Conversely, if we assume that $v_1^* = v_2^*$ and $k_1^* = k_2^*$, then we can deduce that $r_1^* \cdot \lambda_2^* = r_2^* \cdot \lambda_1^*$. This means that the conditions

$$v_1^* = v_2^*, \qquad k_1^* = k_2^* \tag{4.20}$$

are equivalent to the conditions

$$v_1^* = v_2^*, \qquad r_1^* \cdot \lambda_2^* = r_2^* \cdot \lambda_1^*, \tag{4.21}$$

and hence either (4.20) or (4.21) are necessary and sufficient conditions for the design (4.19) to have only two distinct associate classes. Under (4.20) or (4.21) we can further deduce that

$$\frac{b_2^*}{b_1^*} = \frac{r_2^*}{r_1^*} = \frac{\lambda_2^*}{\lambda_1^*}. \tag{4.22}$$

These results are stated in the following corollary.

COROLLARY 4.2.2. *The necessary and sufficient conditions for the Kronecker product* $N = N_{(1)\mathrm{BIB}} \times N_{(2)\mathrm{BIB}}$ *of the* BIB *designs* $N_{(1)\mathrm{BIB}}$ *and* $N_{(2)\mathrm{BIB}}$ *with respective sets of parameters*

$$v_1^*, b_1^*, r_1^*, k_1^*, \lambda_1^*,$$
$$v_2^*, b_2^*, r_2^*, k_2^*, \lambda_2^*,$$

to have only two distinct associate classes are

$$v_1^* = v_2^* = v, \quad \text{say}, \qquad k_1^* = k_2^* = k, \quad \text{say}.$$

If these conditions are satisfied then we have $b_2^*/b_1^* = r_2^*/r_1^* = \lambda_2^*/\lambda_1^* = \mu$, *say, where* μ *is a positive fraction, and in this case the parameters of N are expressed in terms of those of* $N_{(1)\mathrm{BIB}}$ *and* $N_{(2)\mathrm{BIB}}$ *by the equations*

$$\left.\begin{aligned} &v' = v^2, \quad b' = \mu \cdot (b_1^*)^2, \quad r' = \mu \cdot (r_1^*)^2, \quad k' = k^2, \\ &n_1' = 2(v-1), \quad n_2' = (v-1)^2, \quad \lambda_1' = \mu \cdot r_1^* \cdot \lambda_1^*, \quad \lambda_1' = \mu \cdot (\lambda_1^*)^2, \\ &(p_{\nu\mu}'^1) = \begin{pmatrix} v-2 & v-1 \\ v-1 & (v-1)(v-2) \end{pmatrix}, \quad (p_{\nu\mu}'^2) = \begin{pmatrix} 2 & 2(v-2) \\ 2(v-2) & (v-2)^2 \end{pmatrix}. \end{aligned}\right\} \tag{4.23}$$

The following definition of a cyclic design is given by Bose and Shimamoto [9].

Consider a PBIB design $N_{\mathrm{PBIB}}^{(2)}$ with two associate classes and with parameters $v, b, r, k, n_i, \lambda_i, p_{jk}^i, \quad i, j, k = 1, 2$. Let its treatments be designated by the integers $1, 2, \cdots, v$. The design $N_{\mathrm{PBIB}}^{(2)}$ is said to be a C (cyclic) design if the first associates of the treatment i of $N_{\mathrm{PBIB}}^{(2)}$ regarded as a PBIB design are the treatments

$$i + d_1, \quad i + d_2, \quad \cdots, \quad i + d_{n_1} \quad \mod v$$

where the d's satisfy the conditions:

(i) the d's are all different and $0 < d_j < v$ for $j = 1, 2, \cdots, n_1$;

(ii) among the $n_1(n_1 - 1)$ differences $d_j - d_{j'}$, $j, j' = 1, 2, \cdots, n_1$; $j \neq j'$; reduced mod v each of the numbers $d_1, d_2, \cdots, d_{n_1}$ occurs α times whereas each of the numbers $e_1, e_2, \cdots, e_{n_2}$ occurs β times where $d_1 d_2, \cdots, d_{n_1}, e_1, e_2, \cdots, e_{n_2}$ are all the different $v - 1$ numbers $1, 2, \cdots, v - 1$.

Clearly it is necessary that

$$n_1\alpha + n_2\beta = n_1(n_1 - 1). \tag{4.24}$$

The parameters p^i_{jk}, $i, j, k = 1, 2$, of $N^{(2)}_{\text{PBIB}}$ are in this case given by

$$(p^1_{jk}) = \begin{pmatrix} \alpha & n_1 - \alpha - 1 \\ n_1 - \alpha - 1 & n_2 - n_1 + \alpha + 1 \end{pmatrix}, \qquad (p^2_{jk}) = \begin{pmatrix} \beta & n_1 - \beta \\ n_1 - \beta & n_2 - n_1 + \beta + 1 \end{pmatrix}. \tag{4.25}$$

If we take $\alpha = v - 2$ and $\beta = 2$, then we find that the necessary conditions (4.24) and (4.25) are satisfied by the corresponding parameters in (4.23). Let the treatments of the design N of (4.23) be designated by integers $1, 2, \cdots, v'$. Then, according to the method of identification of the various associate classes in N described in the proof of Theorem 4.2, the first associates of the treatment 1 in N are treatments $2, 3, \cdots, v, v + 1, 2v + 1, \cdots, v^2 - v + 1$. The corresponding d's are clearly $1, 2, \cdots, v - 1, v, 2v, \cdots, v^2 - v$. If we form the

$$2(v - 1)(2v - 3)$$

differences $d_j - d'_j$, $j, j' = 1, 2, \cdots, 2(v - 1)$; $j \neq j'$; of these $2(v - 1)$ d's, it is obvious that $d_1 = 1$ will occur in these differences exactly $v - 1$ times whereas if the design N of (4.23) were a cyclic design, $d_1 = 1$ must occur only $\alpha = v - 2$ times. Hence we find that the design N of (4.23) cannot be a cyclic design even though its parameters satisfy the necessary conditions (4.24) and (4.25).

5. Construction of certain PBIB designs. From the results of Section 4 we find that two Kronecker products which give PBIB designs with two associate classes are

(i) $N_2 \times N_{\text{BIB}}$ (Corollary 4.1.1),

(ii) $N_{(1)\text{BIB}} \times N_{(2)\text{BIB}}$ where $N_{(1)\text{BIB}}$ and $N_{(2)\text{BIB}}$ are two BIB designs for which $v^*_1 = v^*_2$ and $k^*_1 = k^*_2$ (Corollary 4.2.2).

Bose, Shrikhande, and Bhattacharya [13] have obtained certain singular GD designs by applying Corollary 4.1.1, which is the only way of getting them. The following example illustrates Corollary 4.2.2.

EXAMPLE 5.1. Let us take $N_{(1)\text{BIB}}$ to be the BIB design defined by the parameters

$$v^* = b^* = 4, \qquad r^* = k^* = 3, \qquad \lambda^* = 2 \tag{5.1}$$

(cf. Cochran and Cox [6]). Let $N_{(2)BIB}$ be the same as $N_{(1)BIB}$ so that the value of μ in Corollary 4.2.2 is 1. Then clearly the parameters v', b', r', k' of

$$N = N_{(1)BIB} \times N_{(2)BIB}$$

are

$$v' = b' = 16, \qquad r' = k' = 9. \tag{5.2}$$

Let the treatments of N be designated by the integers $1, 2, \cdots, 16$. The proof of Theorem 4.2 contains a description of the method of identifying the various associate classes of N. According to this method we get the following identifications.

Treatment	1	2	6
First associates	2, 3, 4, 5, 9, 13.	1, 3, 4, 6, 10, 14.	2, 5, 7, 8, 10, 14.
Second associates	6, 7, 8, 10, 11, 12, 14, 15, 16.	5, 7, 8, 9, 11, 12, 13, 15, 16.	1, 3, 4, 9, 11, 12, 13, 15, 16.

It is clear that the parameters n_1', n_2', λ_1', λ_2' of N are given by

$$n_1' = 6, \qquad n_2' = 9, \qquad \lambda_1' = 6, \qquad \lambda_2' = 4. \tag{5.3}$$

Further the comparisons of the associate classes of treatment 1 with those of treatments 2 and 6 respectively lead to

$$(p_{jk}'^{1}) = \begin{pmatrix} 2 & 3 \\ 3 & 6 \end{pmatrix}, \qquad (p_{jk}'^{2}) = \begin{pmatrix} 2 & 4 \\ 4 & 4 \end{pmatrix} \tag{5.4}$$

where $p_{jk}'^{i}$, $i, j, k = 1, 2$, are parameters of N. It may be noted that the design N is not cyclic even though the necessary conditions (4.24) and (4.25) are satisfied by its parameters.

The equations (5.2) to (5.4) give all the parameters of the design N. The blocks of the design N are shown below.

(1, 2, 3, 5, 6, 7, 9, 10, 11), (1, 2, 3, 5, 6, 7, 13, 14, 15),
(1, 2, 4, 5, 6, 8, 9, 10, 12), (1, 2, 4, 5, 6, 8, 13, 14, 16),
(1, 3, 4, 5, 7, 8, 9, 11, 12), (1, 3, 4, 5, 7, 8, 13, 15, 16),
(2, 3, 4, 6, 7, 8, 10, 11, 12), (2, 3, 4, 6, 7, 8, 14, 15, 16),
(1, 2, 3, 9, 10, 11, 13, 14, 15), (5, 6, 7, 9, 10, 11, 13, 14, 15),
(1, 2, 4, 9, 10, 12, 13, 14, 16), (5, 6, 8, 9, 10, 12, 13, 14, 16),
(1, 3, 4, 9, 11, 12, 13, 15, 16), (5, 7, 8, 9, 11, 12, 13, 15, 16),
(2, 3, 4, 10, 11, 12, 14, 15, 16), (6, 7, 8, 10, 11, 12, 14, 15, 16).

From the remaining results of Section 4 we find that the following Kronecker products give PBIB designs with more than two associate classes.

(i) $N_2 \times N_{PBIB}^{(s)}$ (Theorem 4.1),

(ii) $N_{PBIB}^{(s)} \times N_{PBIB}^{(t)}$ (Theorem 4.2),

(iii) $N_{(1)\text{BIB}} \times N_{(2)\text{BIB}}$ (Corollary 4.2.1).

The following example illustrates Corollary 4.2.1.

EXAMPLE 2. Let $N_{(1)\text{BIB}}$ and $N_{(2)\text{BIB}}$ be the BIB designs defined by the sets of parameters

$$v_1^* = b_1^* = 3, \quad r_1^* = k_1^* = 2, \quad \lambda_1^* = 1, \tag{5.5}$$

$$v_2^* = 5, \quad b_2^* = 10, \quad r_2^* = 4, \quad k_2^* = 2, \quad \lambda_2^* = 1 \tag{5.6}$$

respectively (cf. Cochran and Cox [6]). Then clearly the parameters v', b', r', k' of $N = N_{(1)\text{BIB}} \times N_{(2)\text{BIB}}$ are given by

$$v' = 15, \quad b' = 30, \quad r' = 8, \quad k' = 4. \tag{5.7}$$

Let the treatments of N be designated by integers $1, 2, \cdots, 15$. According to the method of identifying the various associate classes of N described in the proof of Theorem 4.2, we get the following identifications.

Treatment	1	2	6	7
First associates	2, 3, 4, 5.	1, 3, 4, 5.	7, 8, 9, 10.	6, 8, 9, 10.
Second associates	6, 11.	7, 12.	1, 11.	2, 12.
Third associates	7, 8, 9, 10, 12, 13, 14, 15.	6, 8, 9, 10, 11, 13, 14, 15.	2, 3, 4, 5, 12, 13, 14, 15.	1, 3, 4, 5, 11, 13, 14, 15.

Also it is clear that the parameters n_1', n_2', n_3', λ_1', λ_2', λ_3' of N are given by

$$n_1' = 4, \quad n_2' = 2, \quad n_3' = 8, \quad \lambda_1' = 4, \quad \lambda_2' = 2, \quad \lambda_3' = 1. \tag{5.8}$$

Further, the comparisons of the associate classes of treatment 1 with those of treatments 2, 6, and 7 respectively lead to

$$(p_{jk}'^{1}) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 6 \end{pmatrix}, \quad (p_{jk}'^{2}) = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 4 \end{pmatrix}, \quad (p_{jk}'^{3}) = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \\ 3 & 1 & 3 \end{pmatrix} \tag{5.9}$$

where $p_{jk}'^{i}$, $i, j, k = 1, 2, 3$, are parameters of N.

The equations (5.7) to (5.9) give all the parameters of the design N. The blocks of the design are as shown below.

(1, 2, 6, 7),	(1, 3, 6, 8),	(1, 4, 6, 9),	(1, 5, 6, 10),	(2, 3, 7, 8),
(2, 4, 7, 9),	(2, 5, 7, 10),	(3, 4, 8, 9),	(3, 5, 8, 10),	(4, 5, 9, 10),
(1, 2, 11, 12),	(1, 3, 11, 13),	(1, 4, 11, 14),	(1, 5, 11, 15),	(2, 3, 12, 13),
(2, 4, 12, 14),	(2, 5, 12, 15),	(3, 4, 13, 14),	(3, 5, 13, 15),	(4, 5, 14, 15),
(6, 7, 11, 12),	(6, 8, 11, 13),	(6, 9, 11, 14),	(6, 10, 11, 15),	(7, 8, 12, 13),
(7, 9, 12, 14),	(7, 10, 12, 15),	(8, 9, 13, 14),	(8, 10, 13, 15),	(9, 10, 14, 15).

Example 5.3. Consider the PBIB design N with three associate classes and with parameters:

$$v = b = pq, \qquad r = k = p + q - 1,$$

$$(5.10) \qquad n_1 = p - 1, \qquad n_2 = q - 1, \qquad n_3 = (p-1)(q-1),$$

$$\lambda_1 = p, \qquad \lambda_2 = q, \qquad \lambda_3 = 2,$$

$$(p_{jk}^1) = \begin{bmatrix} p-2 & 0 & 0 \\ 0 & 0 & q-1 \\ 0 & q-1 & (p-2)(q-1) \end{bmatrix},$$

$$(p_{jk}^2) = \begin{bmatrix} 0 & 0 & p-1 \\ 0 & q-2 & 0 \\ p-1 & 0 & (p-1)(q-2) \end{bmatrix},$$

$$(p_{jk}^3) = \begin{bmatrix} 0 & 1 & p-2 \\ 1 & 0 & q-2 \\ p-2 & q-2 & (p-2)(q-2) \end{bmatrix}$$

where p, q are positive integers $\geqq 2$ and $j, k = 1, 2, 3$.

This design taken from Bose and Nair [2] very much resembles the Kronecker product of two BIB designs. Let us suppose that the above design is the Kronecker product of the two BIB designs defined by the sets of parameters

$$v_1^* = q, \qquad b_1^*, \qquad r_1^*, \qquad k_1^*, \qquad \lambda_1^*$$

and

$$v_2^* = p, \qquad b_2^*, \qquad r_2^*, \qquad k_2^*, \qquad \lambda_2^*.$$

From Corollary 4.2.1 it follows that we must have

$$\lambda_1 = r_1^*\lambda_2^* = p, \qquad \lambda_2 = r_2^*\lambda_1^* = q, \qquad \lambda_3 = \lambda_1^* \cdot \lambda_2^* = 2.$$

Hence

$$pq = \lambda_1 \cdot \lambda_2 = \lambda_1^*\lambda_2^*r_1^*r_2^* = 2r_1^*r_2^* = 2k_1^*k_2^* = 2(p + q - 1),$$

which leads to

$$(5.11) \qquad (p-2)(q-2) = 2.$$

This means that a necessary condition for the design N of (5.10) to be the Kronecker product of two BIB designs is (5.11).

Also since p and q are positive integers, we must have from (5.11) either $p = 3$ and $q = 4$ or $p = 4$ and $q = 3$. It is enough to consider one case, say, $p = 4$, $q = 3$. With these values it is clear that (5.11) is satisfied. The corresponding

BIB designs are defined by the sets of parameters

$$v_1^* = b_1^* = 3, \quad r_2^* = k_2^* = 2, \quad \lambda_1^* = 1,$$

and

$$v_2^* = b_2^* = 4, \quad r_2^* = k_2^* = 3, \quad \lambda_2^* = 2.$$

Now applying Corollary 4.2.1 to these two BIB designs it is easily verified that their Kronecker product has the parameters of design N of (5.10) for which $p = 4$ and $q = 3$. Thus we find that the condition (5.11) is also sufficient for the design N of (5.10) to be constructible as the Kronecker product of two BIB designs.

It has been remarked by Bose and Nair [2] that the design N of (5.10) with three associate classes reduces to a PBIB design with two associate classes if $p = 2$ or $q = 2$. Because of Lemma 4.1 we can further add that the design N with three associate classes reduces to a PBIB design with two associate classes if $p = q$.

The method of taking Kronecker product of designs has been used to prove the impossibility of a certain class of PBIB designs and to analyse some other class of designs. It is hoped to publish at a later date some results in this direction.

I wish to express my sincere thanks to Professor M. C. Chakrabarti under whose guidance this work was carried out.

REFERENCES

[1] F. D. MURNAGHAN, *The Theory of Group Representations*, The Johns Hopkins Press, Baltimore, 1938, pp. 69–70.

[2] R. C. BOSE AND K. R. NAIR, "Partially balanced incomplete block designs", *Sankyā*, Vol. 4 (1939), pp. 337–372.

[3] C. R. RAO, "General methods of analysis for incomplete block designs", *J. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 541–561.

[4] R. C. BOSE, "A note on Fisher's inequality for balanced incomplete block designs", *Ann. Math. Stat.*, Vol. 20 (1949), pp. 619–620.

[5] SCHUTZENBERGER, "A non-existence theorem for an infinite family of symmetrical block designs", *Ann. of Eugenics*, Vol. 14 (1949), pp. 286–287.

[6] W. G. COCHRAN AND G. M. COX, *Experimental Designs*, John Wiley and Sons, New York, 1950, p. 327.

[7] S. S. SHRIKHANDE, "The impossibility of certain symmetrical balanced incomplete block designs", *Ann. Math. Stat.*, Vol. 21 (1950), pp. 106–111.

[8] R. C. BOSE AND W. S. CONNOR: "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[9] R. C. BOSE AND T. SHIMAMOTO, "Classification and analysis of partially balanced incomplete block designs with two associate classes", *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[10] W. S. CONNOR, "On the structure of balanced incomplete block designs", *Ann. Math. Stat.*, Vol. 23 (1952) pp. 57–71.

[11] S. S. SHRIKHANDE," On the dual of some balanced incomplete block designs", *Biometrics* Vol. 8 (1952), pp. 66–72.

[12] S. S. SHRIKHANDE, "The non-existence of certain affine resolvable balanced incomplete block designs", *Canadian J. Math.*, Vol. 5 (1953), p. 413–420.

[13] R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs", *Ann. Math. Stat.*, Vol. 24 (1953), pp. 167–195.

[14] P. M. Roy, "A note on the relation between BIB and PBIB designs", *Science and Culture*, Vol. 19 (1953), pp. 40–41.

[15] P. M. Roy, "A note on the method of inversion of statistical designs", *Science and Culture*, Vol. 19 (1953), pp. 440–441.

[16] W. S. Connor and W. H. Clatworthy, "Some theorems for partially balanced designs", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–112.

[17] M. Zelen, "A note on partially balanced designs", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 599–602.

BOUNDS FOR THE DISTRIBUTION FUNCTION OF A SUM OF INDEPENDENT, IDENTICALLY DISTRIBUTED RANDOM VARIABLES[1]

By Wassily Hoeffding and S. S. Shrikhande

University of North Carolina and College of Science, Nagpur, India

Summary. The problem is considered of obtaining bounds for the (cumulative) distribution function of the sum of n independent, identically distributed random variables with k prescribed moments and given range. For $n = 2$ it is shown that the best bounds are attained or arbitrarily closely approached with discrete random variables which take on at most $2k + 2$ values. For nonnegative random variables with given mean, explicit bounds are obtained when $n = 2$; for arbitrary values of n, bounds are given which are asymptotically best in the "tail" of the distribution. Some of the results contribute to the more general problem of obtaining bounds for the expected value of a given function of independent, identically distributed random variables when the expected values of certain functions of the individual variables are given. Although the results are modest in scope, the authors hope that this paper will draw attention to a problem of both mathematical and statistical interest.

1. Introduction. This paper considers part of the following general problem. Let $\mathfrak{D}$ be the class of all dfs (distribution functions) $F(x)$ on the real line which satisfy the conditions

$$\int g_i(x)\, dF(x) = c_i, \quad i = 1, \cdots, k; \qquad F(x) = \begin{cases} 0 & x < A, \\ 1 & x > B, \end{cases}$$

where the functions $g_1(x), \cdots, g_k(x)$ and the constants $c_1, \cdots, c_k$, A, and B are given. We allow that $A = -\infty$ and/or $B = \infty$. Here and in what follows, when the domain of integration is not indicated, the integral extends over the entire range of the variables involved.

Let $K(x_1, \cdots, x_n)$ be a function such that

$$\psi(F) = \int \cdots \int K(x_1, \cdots, x_n)\, dF(x_1) \cdots dF(x_n)$$

exists for all F in $\mathfrak{D}$ in the sense that the multiple integral is equal to the repeated integral taken in an arbitrary order. The problem is to determine upper and lower bounds for $\psi(F)$ when F is in $\mathfrak{D}$.

For $n = 1$, $g_i(x) = x^i$, and $K(x) = 1$ or 0 according as $x \leq t$ or $> t$, as well

Received December 10, 1953.

[1] This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

as for other functions $K(x)$, an extensive literature on the subject exists; for some references see [4].

For n arbitrary, Robbins [6] showed that the Bienaymé-Tchebycheff bound for $\Pr(|X_1 + \cdots + X_n| \geqq t)$, where the X_i are independent and identically distributed with zero mean and given variance, can be improved when $n > 1$. Plackett [5], Gumbel [2], and Hartley and David [3] obtained the best possible bounds for the expected sample range and the expected value of the largest observation, in the case when the mean and the variance are given, assuming that the common df is continuous. In a problem analogous to the general problem stated above, but without the assumption that the n variables are identically distributed, one of the authors [4] showed that under general conditions the best bounds are attained or arbitrarily closely approached with step-functions in $\mathfrak{D}$ which have at most $k + 1$ steps.

The present paper concentrates attention on the case where $K = 1$ or 0 according as a given function $f(x_1, \cdots, x_n)$ is or is not contained in a given set. The method used permits one to obtain the closest bounds only for $n = 2$. If n is even, $f = x_1 + \cdots + x_n$, and $g_i(x) = x^i$, the bounds for $n = 2$ can be applied in an obvious way, but in general will not be the best ones. More general functions K are considered only insofar as they can be handled by the same method.

Theorem 2.1 states conditions under which we need consider only step-functions in $\mathfrak{D}$. Theorems 2.2 and 2.3 show that for functions $K(x, y)$ of a certain type we may restrict our attention to step-functions with a bounded number of steps. In Theorem 3.1 an explicit expression for the least upper bound of $\Pr(X + Y \geqq t)$ is obtained when X and Y are nonnegative, independent, and identically distributed with given mean. In Section 4 bounds for the analogous case with n summands are considered.

2. The least upper bound of $\iint K(x, y)\, dF(x)\, dF(y)$. Let $K(x, y)$ be a function such that

$$\psi(F) = \iint K(x, y)\, dF(x)\, dF(y) \tag{2.1}$$

exists for all F in $\mathfrak{D}$, in the sense that the double integral equals the repeated integral. The problem is to determine the least upper bound of $\psi(F)$ for all F in $\mathfrak{D}$.

Let $\mathfrak{D}^*$ be the class of all F in $\mathfrak{D}$ which are step-functions with a finite number of steps. The following theorem shows that if $\mathfrak{D}$ is the class of dfs with k prescribed moments and given range, and $\psi(F)$ is the probability that two independent observations on a random variable with df F fall into a set of a rather general type, we may confine our attention to dfs in $\mathfrak{D}^*$.

THEOREM 2.1. *Let $g_i(x) = x^{m_i}$, where $m_1, \cdots, m_k$ are positive integers. Let $K(x, y) = 1$ or 0 according as (x, y) is or is not contained in a Borel set S such that the sets $\{x: (x, y) \in S, y \text{ fixed}\}$ and $\{y: (x, y) \in S, x \text{ fixed}\}$ are unions of a*

finite and bounded number of intervals (which may be infinite). Then

$$\sup_{F \in \mathfrak{D}} \psi(F) = \sup_{F \in \mathfrak{D}^*} \psi(F).$$

The theorem follows immediately from an obvious analog of Lemma 2.1 in [4] and Lemma 3.1 and Theorem 4.1 of [4].

It can be seen from [4] that the reduction to distributions in $\mathfrak{D}^*$ is possible under more general conditions.

We shall now derive sufficient conditions under which, given a step-function F in $\mathfrak{D}$ with m steps, we can construct a step-function G in $\mathfrak{D}$ with less than m steps such that $\psi(G) \geqq \psi(F)$.

A step-function F in $\mathfrak{D}$ with exactly m steps is of the form

$$F(x) = P_j \quad \text{if} \quad a_j \leqq x < a_{j+1}, \qquad j = 0, 1, \cdots, m, \tag{2.2}$$

where

$$-\infty = a_0 < a_1 < a_2 < \cdots < a_m < a_{m+1} = \infty, \qquad A \leqq a_1, \qquad a_m \leqq B; \tag{2.3}$$

$$0 = P_0 < P_1 < \cdots < P_{m-1} < P_m = 1; \tag{2.4}$$

$$\sum_{j=1}^{m-1} h_{ij} P_j = c_i - g_i(a_m), \qquad i = 1, \cdots, k; \tag{2.5}$$

$$h_{ij} = g_i(a_j) - g_i(a_{j+1}), \qquad i = 1, \cdots, k; \; j = 1, \cdots, m-1. \tag{2.6}$$

Let

$$G(x) = P_j + t\,D_j \quad \text{if} \quad a_j \leqq x < a_{j+1}, \qquad j = 0, \cdots, m. \tag{2.7}$$

In order that $G(x)$ be a df in $\mathfrak{D}$ it is sufficient that the numbers t and D_j satisfy the conditions

$$D_0 = D_m = 0; \tag{2.8}$$

$$0 \leqq P_1 + t\,D_1 \leqq P_2 + t\,D_2 \leqq \cdots \leqq P_{m-1} + t\,D_{m-1} \leqq 1; \tag{2.9}$$

$$\sum_{j=1}^{m-1} h_{ij} D_j = 0, \qquad i = 1, \cdots, k. \tag{2.10}$$

If F and G are defined by (2.2) and (2.7), we have

$$\psi(G) - \psi(F) = t \sum_{j=1}^{m-1} L_j D_j + t^2 \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} L_{ij} D_i D_j, \tag{2.11}$$

where, with $K_{ij} = K(a_i, a_j)$,

$$L_j = \sum_{i=1}^{m} (K_{ij} + K_{ji} - K_{i,j+1} - K_{j+1,i})(P_i - P_{i-1}), \tag{2.12}$$

$$L_{ij} = K_{ij} - K_{i+1,j} - K_{i,j+1} + K_{i+1,j+1}. \tag{2.13}$$

LEMMA 2.1. *Let F be a step-function in $\mathfrak{D}$ with exactly m steps, defined by* (2.2)

to (2.6), where $m > k + 1$. Suppose that the integers $u_1, \cdots, u_{k+1}$ can be so chosen that

$$1 \leqq u_1 < u_2 < \cdots < u_{k+1} \leqq m - 1$$

and the equations

$$\sum_{r=1}^{k+1} h_{iu_r} x_r = 0, \qquad i = 1, \cdots, k, \tag{2.14}$$

imply

$$\sum_{r=1}^{k+1} \sum_{s=1}^{k+1} L_{u_r u_s} x_r x_s \geqq 0. \tag{2.15}$$

Then there exists a step-function G in $\mathfrak{D}$ with less than m steps, for which $\psi(G) \geqq \psi(F)$.

Proof. Let $G(x)$ be defined by (2.7), and let $D_j = 0$ for $j \neq u_1, \cdots, u_{k+1}$. Let $\lambda = 1$ or 0 according as the rank of the matrix

$$\begin{Vmatrix} h_{1u_1} & \cdots & h_{1u_{k+1}} \\ \cdots & \cdots & \cdots \\ h_{ku_1} & \cdots & k_{ku_{k+1}} \\ L_{u_1} & \cdots & L_{u_{k+1}} \end{Vmatrix}$$

is equal to or less than $k + 1$. Then the equations (2.14) and

$$\sum_{r=1}^{k+1} L_{u_r} x_r = \lambda$$

have a solution $(D_{u_1}, \cdots, D_{u_{k+1}}) \neq (0, \cdots, 0)$. Having thus fixed the D_j, let t be the largest number which satisfies the inequalities (2.9). This number exists and is positive. With this choice of the numbers t and D_j, G is a step-function in $\mathfrak{D}$ with less than m steps. Furthermore, by (2.11),

$$\psi(G) - \psi(F) = t\lambda + t^2 \sum_{r=1}^{k+1} \sum_{s=1}^{k+1} L_{u_r u_s} D_{u_r} D_{u_s} \geqq 0.$$

The proof is complete.

The next theorem shows that if $K(x, y)$ is of a certain form, and if we restrict ourselves to the class $\mathfrak{D}^*$ of step-functions in $\mathfrak{D}$ with a finite number of steps, we need consider only step-functions with a bounded number of steps.

Let $\mathfrak{D}_m$ be the class of all F in $\mathfrak{D}$ which are step-functions with at most m steps.

Theorem 2.2. *Suppose that $K(x, y)$ is of the form*

$$K(x, y) = \sum_{i=0}^{k} \sum_{j=0}^{k} a_{tij} g_i(x) g_j(y) \qquad \text{if } b_{t-1} \leqq f(x, y) < b_t, \qquad t = 1, \cdots, s,$$

where $g_0(x) = 1$, the a_{tij} are arbitrary constants, the b_t satisfy

$$-\infty = b_0 < b_1 < \cdots < b_{s-1} < b_s = \infty,$$

and $f(x, y)$ is a strictly increasing function in each of its arguments when the other argument is fixed. Then

$$\sup_{F \varepsilon \mathfrak{D}^*} \psi(F) = \sup_{F \varepsilon \mathfrak{D}_{sk+s}} \psi(F).$$

The theorem remains true if in the inequalities $b_{t-1} \leqq f(x, y) < b_t$ some signs $\leqq$ are replaced by $<$ or vice versa, provided that the s sets defined by the inequalities cover the entire plane.

PROOF. Let $F(x)$, as defined by (2.2) to (2.6), be an arbitrary step-function in $\mathfrak{D}$ with exactly m steps, where $m > sk + s$. It is sufficient to construct a step-function G in $\mathfrak{D}$ with less than m steps such that $\psi(G) \geqq \psi(F)$. Let m_t, for $t = 1, \cdots, s$, denote the number of indices u, with $1 \leqq u \leqq m$, for which

$$b_{t-1} \leqq f(a_u, a_u) < b_t.$$

Then $s \max (m_t) \geqq (m_1 + \cdots + m_s) = m > s(k + 1)$. Hence there exists a t for which $m_t \geqq k + 2$ and an integer n such that

$$b_{t-1} \leqq f(a_n, a_n) < f(a_{n+k+1}, a_{n+k+1}) < b_t.$$

The assumption about $f(x, y)$ implies that

$$K_{vw} = \sum_{i=0}^{k} \sum_{j=0}^{k} a_{tij} g_i(a_v) g_j(a_w) \qquad n \leqq v, w \leqq n + k + 1.$$

By (2.13) and (2.6) this implies

$$L_{vw} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{tij} h_{iu} h_{jw} \qquad n \leqq v, w \leqq n + k.$$

Hence if we let $u_r = n + r - 1$ for $r = 1, 2, \cdots, k + 1$, the conditions of Lemma 2.1 are satisfied. The proof is complete.

If $g_i(x) = x^i$, that is, if $\mathfrak{D}$ is the class of distributions with given moments up to order k and given range, the assumption of Theorem 2.2 means that $K(x, y)$ is piecewise polynomial, of bounded degrees, in sections of the plane separated by curves of *negative* slope. If $K(x, y)$ is piecewise polynomial in sections separated by curves of *positive* slope, a similar reduction of the problem to the case of step-functions with a bounded number of steps is in general impossible. For example, let $K(x, y) = \max (x, y)$, and let $\mathfrak{D}$ be the class of dfs F with

$$\int x \, dF(x) = 0, \qquad \int x^2 \, dF(x) = 1.$$

Under the restriction to continuous functions $F(x)$, this is a special case of a problem considered by Hartley and David [3] and Gumbel [2]. For an arbitrary df $F(x)$ we can write

$$\psi(F) = 2 \int x\bar{F}(x) \, dF(x), \qquad \bar{F}(x) = \tfrac{1}{2}[F(x - 0) + F(x + 0)].$$

Using Schwarz's inequality, we have for any constant c and any F in $\mathfrak{D}$

$$\psi(F) + c = 2 \int (x + c)\bar{F}(x)\, dF(x) \leqq 2 \left((1 + c^2) \int \bar{F}(x)^2\, dF(x) \right)^{1/2}.$$

If $F(x)$ is continuous, $\int \bar{F}(x)^2\, dF(x) = \frac{1}{3}$, and the bound

$$\psi(F) \leqq \min_c \{2 \cdot 3^{-\frac{1}{2}}(1 + c^2)^{\frac{1}{2}} - c\}$$

is attained with a continuous df in $\mathfrak{D}$, as shown by Hartley and David.

Now let $F(x)$ be a step-function with at most m steps which takes on the values $0 = P_0 \leqq P_1 \leqq \cdots \leqq P_{m-1} \leqq P_m = 1$. Then

$$4 \int \bar{F}(x)^2\, dF(x) = \sum_{j=1}^{m} (P_{j-1} + P_j)^2 (P_j - P_{j-1}).$$

This can be written

$$12 \int \bar{F}(x)^2\, dF(x) = 4 - \sum_{j=1}^{m} p_j^3, \qquad p_j = P_j - P_{j-1}.$$

The conditions $\sum p_j = 1$ and $p_j \geqq 0$ imply $\sum p_j^3 \geqq m^{-2}$. Hence

$$\int \bar{F}(x)^2\, dF(x) \leqq \tfrac{1}{3} - 1/12m^2,$$

and the Hartley-David bound cannot be approached arbitrarily closely with a step-function in $\mathfrak{D}$ having a bounded number of steps.

Combining Theorems 2.1 and 2.2 we can state that if the conditions of both theorems are satisfied, then

$$\sup_{F \in \mathfrak{D}} \psi(F) = \sup_{F \in \mathfrak{D}_{2k+2}} \psi(F).$$

In particular, the conditions of Theorem 2.2 are fulfilled if $\psi(F) = P_F\{f(X, Y) \geqq c\}$, or $= P_F\{\,|f(X, Y)| \geqq c\}$, etc., where $P_F\{\cdots\}$ is the probability of the event in braces when X and Y are independent with common df F, and $f(x, y)$ has the property stated in the theorem. Using Theorem 2.1, we obtain:

THEOREM 2.3. *Let $\mathfrak{D}$ be the class of dfs $F(x)$ which satisfy the conditions*

$$\int x^{m_i}\, dF(x) = c_i, \quad i = 1, \cdots, k, \qquad F(x) = \begin{cases} 0 & x < A, \\ 1 & x > B, \end{cases}$$

with given integers $m_1, \cdots, m_k$ and given numbers $c_1, \cdots, c_k, A, B$, where we may have $A = -\infty$ and/or $B = \infty$. Let $f(x, y)$ be a strictly increasing function in each of its arguments when the other argument is fixed. Then

$$\sup_{F \in \mathfrak{D}} P_F\{f(X, Y) \geqq c\} = \sup_{F \in \mathfrak{D}_{2k+2}} P_F\{f(X, Y) \geqq c\}.$$

3. The least upper bound of $P(X + Y \geqq t)$ when X and Y are nonnegative, independent, and identically distributed with given mean. As an application of the results of Section 2 we shall prove the following theorem.

THEOREM 3.1. *Let X and Y be two independent random variables with common* cdf $F(x)$. *Let $\mathfrak{D}$ be the class of* dfs F *with $F(x) = 0$ for $x < 0$ and $\int x\,dF(x) = \mu$, where $\mu > 0$. Then*

$$(3.1)\qquad \sup_{F\in\mathfrak{D}} P_F(X + Y \geqq c\mu) = \begin{cases} 1, & c \leqq 2; \\ 4/c^2, & 2 \leqq c \leqq \frac{5}{2}; \\ 2/c - 1/c^2, & \frac{5}{2} \leqq c. \end{cases}$$

The three bounds are attained with the respective distributions

$$P(X = \mu) = 1;$$

$$P(X = 0) = 1 - 2/c, \qquad P(X = \tfrac{1}{2}c\mu) = 2/c;$$

$$P(X = 0) = 1 - 1/c, \qquad P(X = c\mu) = 1/c.$$

Theorem 3.1 should be compared with the solution by Birnbaum, Raymond, and Zuckerman [1] of the analogous problem without the restriction that X and Y be identically distributed. If $M(c)$ denotes the least upper bound of $P(X + Y \geqq c\mu)$ when X and Y are nonnegative, independent, and have the common mean μ, we have by [1]

$$(3.2)\qquad M(c) = \begin{cases} 1, & c \leqq 2; \\ 1/(c - 1), & 2 \leqq c \leqq \frac{1}{2}(3 + \sqrt{5}); \\ 2/c - 1/c^2, & \frac{1}{2}(3 + \sqrt{5}) \leqq c. \end{cases}$$

Hence the bound (3.1) is smaller than the Birnbaum-Raymond-Zuckerman bound if and only if $2 < c < \frac{1}{2}(3 + \sqrt{5})$.

PROOF OF THEOREM 3.1. We may and shall assume that $\mu = 1$. By Theorem 2.3 we need consider only dfs F in $\mathfrak{D}$ which are step-functions with $m \leqq 4$ steps. Then F is of the form (2.2) to (2.4), where $A = 0$ and $B = \infty$, and $\sum_1^m a_j(P_j - P_{j-1}) = 1$. We have

$$K(x, y) = \begin{cases} 1, & x + y \geqq c; \\ 0, & x + y < c. \end{cases}$$

Hence the numbers $K_{ij} = K(a_i, a_j)$ satisfy the conditions

$$K_{ij} = 0 \text{ or } 1; \qquad K_{ij} = K_{ji}; \qquad K_{ij} \leqq K_{i'j} \quad \text{if} \quad i < i'.$$

The sequence $(K_{11}, K_{22}, \cdots, K_{mm})$ consists of a sequence of zeros followed by a sequency of ones. The reasoning used in the proof of Theorem 2.2 shows that any distribution for which there are more than two consecutive zeros or more than two consecutive ones in this sequence can be replaced by a distribution with less than m steps which does not decrease the value of $\psi(F)$.

Hence for $m = 4$ we need consider only matrices $\|K_{ij}\|$ of the four types

$$\begin{Vmatrix} 0 & 0 & 0 & \cdot \\ 0 & 0 & 0 & \cdot \\ 0 & 0 & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{Vmatrix},$$

where the numbers represented by dots need not be specified. The corresponding matrices $\|L_{ij}\|$ are

$$\begin{array}{cccc} \text{I} & \text{II} & \text{III} & \text{IV} \\ \begin{Vmatrix} 0 & 0 & \cdot \\ 0 & 1 & \cdot \\ \cdot & \cdot & 0 \end{Vmatrix}, & \begin{Vmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{Vmatrix}, & \begin{Vmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \end{Vmatrix}, & \begin{Vmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{Vmatrix}. \end{array}$$

We shall apply Lemma 2.1 to show that in every case there exists a df in $\mathfrak{D}$ with at most three steps which does not decrease the value of $\psi(F)$. It is sufficient to find integers u and v with $1 \leqq u < v \leqq 3$ such that the equation

$$(a_u - a_{u+1})x + (a_v - a_{v+1})y = 0 \tag{3.3}$$

implies

$$L_{uu}x^2 + 2L_{uv}xy + L_{vv}y^2 \geqq 0. \tag{3.4}$$

Inequality (3.4) is satisfied in Case I with $u = 1$ and $v = 2$, and in Cases II and IV with $u = 1$ and $v = 3$. In Case III, when $u = 1$ and $v = 3$, the left side of (3.4) is $-2xy$, which is nonnegative by (3.3), since $a_j - a_{j+1} < 0$.

Hence we may confine our attention to step-functions in $\mathfrak{D}$ with $m \leqq 3$ steps.

If $m = 3$, we have to consider the matrices $\|K_{ij}\|$ of the forms

$$\begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{Vmatrix},$$

$$\begin{Vmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{Vmatrix}.$$

The corresponding matrices $\|L_{ij}\|$ are

$$\begin{array}{cccccc} A & B & C & D & E & F \\ \begin{Vmatrix} 0 & 0 \\ 0 & 1 \end{Vmatrix}, & \begin{Vmatrix} 0 & 1 \\ 1 & -1 \end{Vmatrix}, & \begin{Vmatrix} 0 & 0 \\ 0 & -1 \end{Vmatrix}, & \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix}, & \begin{Vmatrix} 1 & -1 \\ -1 & 0 \end{Vmatrix}, & \begin{Vmatrix} -1 & 0 \\ 0 & 0 \end{Vmatrix}. \end{array}$$

In applying Lemma 2.1 we have to take $u = 1$ and $v = 2$ and show that (3.3) implies (3.4). This is true for the matrices A, D, and E. In Cases B, C, and F, Lemma 2.1 is not applicable.

In Case C, $\psi(F) = 1 - P_2^2$. If $G(x)$ is defined by (2.7) with $m = 3$, we have

$$\psi(G) - \psi(F) = -t\,D_2[2(P_2 + t\,D_2) - t\,D_2],$$

where t and D_2 satisfy the conditions

$$(a_1 - a_2)\,D_1 + (a_2 - a_3)\,D_2 = 0, \tag{3.5}$$

$$0 \leqq P_1 + t\,D_1 \leqq P_2 + t\,D_2 \leqq 1. \tag{3.6}$$

Let $D_2 = -1$. Then D_1 is given by (3.5). Let t be the largest number which satisfies (3.6). Then $t > 0$, G is in $\mathfrak{D}_2$, and $\psi(G) - \psi(F) \geqq 0$.

In case F, $\psi(F) = 1 - P_1^2$, and similar reasoning shows that this case also can be reduced to a step-function with at most two steps.

The only remaining case with $m = 3$ is Case B. Here we can write $\psi(F) = 2p_2p_3 + p_3^2$, where (admitting the possibility that F has less than three steps)

$$p_1 + p_2 + p_3 = 1, \qquad a_1p_1 + a_2p_2 + a_3p_3 = 1; \tag{3.7}$$

$$p_1 \geqq 0, \qquad p_2 \geqq 0, \qquad p_3 \geqq 0; \tag{3.8}$$

$$0 \leqq a_1 \leqq a_2 \leqq a_3; \tag{3.9}$$

$$a_1 + a_3 < c, \qquad 2a_2 < c, \qquad c \leqq a_2 + a_3. \tag{3.10}$$

Expressing $\psi(F)$ in terms of a_1, a_2, a_3, p_1, we get

$$\psi(F) = (1 - p_1)^2 - p_2^2, \qquad p_2 = \frac{a_3 - 1 - (a_3 - a_1)p_1}{a_3 - a_2}.$$

If a_2, a_3, and p_1 are held fixed, $\psi(F)$ is a decreasing function of a_1. Hence we maximize $\psi(F)$ by choosing the least possible value for a_1. This is the greatest of the bounds given by the inequalities $p_j \geqq 0$ and $a_1 \geqq 0$. If this bound is given by one of the equations $p_j = 0$, we get a distribution in $\mathfrak{D}_2$. Hence we may assume that the least value is $a_1 = 0$, so

$$p_2 = \frac{a_3 - 1 - a_3p_1}{a_3 - a_2}$$

and $\psi(F)$ is a decreasing function of a_2 when a_3 and p_1 are fixed. The only lower bound for a_2 which does not necessarily correspond to a distribution in $\mathfrak{D}_2$ is $a_2 = c - a_3$. In this case

$$p_2 = \frac{(1 - p_1)a_3 - 1}{2a_3 - c} = \frac{1 - p_1}{2} + \frac{c(1 - p_1) - 2}{2(2a_3 - c)}.$$

This is a monotonic function of a_3 (possibly a constant) when p_1 is held fixed. Hence the maximum is attained at one of the endpoints of the range of a_3. This range is given by the inequalities (3.8) to (3.10) with $a_1 = 0$ and $a_2 = c - a_3$. Its endpoints correspond either to distributions in $\mathfrak{D}_2$ or (if given by $a_1 + a_3 = c$ or $2a_2 = c$) to cases where the value of $\psi(F)$ exceeds $2p_2p_3 + p_3^2$ and which already have been disposed of.

Thus we need consider only dfs in $\mathfrak{D}_2$.

If $c \leqq 2$, we have $\psi(F) = 1$ for the df in $\mathfrak{D}_1$ which has a single step at $x = 1$. Thus

$$\sup_{F \in \mathfrak{D}} \psi(F) = 1 \qquad c \leqq 2. \tag{3.11}$$

Henceforth we assume that $c > 2$.

A distribution F in $\mathfrak{D}_2$ assigns to the points a_1 and a_2 the respective probabilities

$$p_1 = \frac{a_2 - 1}{a_2 - a_1}, \qquad p_2 = \frac{1 - a_1}{a_2 - a_1}, \qquad 0 \leqq a_1 \leqq 1 \leqq a_2.$$

If $c \leqq 2a_1$, we have $c \leqq 2$, a case already considered. If $c > 2a_2$, then $\psi(F) = 0$, a case which may be disregarded. We are left with the two cases

$$\text{(i)} \quad 2a_1 < c \leqq a_1 + a_2, \qquad \text{(ii)} \quad a_1 + a_2 < c \leqq 2a_2.$$

In Case (i), $\psi(F) = 1 - p_1^2$, which is a decreasing function of a_1. The lower bound for a_1 is max $(0, \quad c - a_2)$.

If $a_1 = 0 \geqq c - a_2$, then $p_1 = 1 - 1/a_2$, so that $\psi(F)$ is a decreasing function of a_2. The lower bound for a_2 is max $(1, c) = c$, and we obtain

$$\psi(F) = 1 - (1 - 1/c)^2 = 2/c - 1/c^2.$$

If $a_1 = c - a_2 \geqq 0$, then

$$p_1 = \frac{a_2 - 1}{2a_2 - c} = \frac{1}{2} + \frac{c - 2}{2(2a_2 - c)},$$

so that $\psi(F)$ is an increasing function of a_2. Since $a_2 \leqq c$, we obtain the same maximum of $\psi(F)$ as in the previous case.

In Case (ii), $\psi(F) = p_2^2$, which is a decreasing function of a_2. Hence we let $a_2 = \frac{1}{2}c$. Then $\psi(F)$ is a decreasing function of a_1, and hence is maximized for $a_1 = 0$. We get $\psi(F) = 4/c^2$. Hence

$$\sup_{F \in \mathfrak{D}} \psi(F) = \max \{2/c - 1/c^2, \; 4/c^2\}, \qquad c > 2. \tag{3.12}$$

Theorem 3.1 now follows from (3.11), (3.12) and the stated conditions under which the bounds are attained.

4. Bounds for $P(X_1 + \cdots + X_n \geqq c)$. Let $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$, and let $\omega_n(t)$ denote the least upper bound of $P(\bar{X}_n \geqq t\mu)$ when $X_1, \cdots, X_n$ are nonnegative, independent, and identically distributed with mean μ. It is easily seen that for every n

$$\omega_n(t) = 1, \quad \text{if } t \leqq 1; \qquad \omega_{sn}(t) \leqq \omega_s(t), \quad s = 1, 2, \cdots.$$

By Markov's inequality, $\omega_1(t) = 1/t$ if $1 \leqq t$. By Theorem 3.1,

$$\omega_2(t) = \begin{cases} 1/t^2 & 1 \leqq t \leqq 5/4, \\ 1/t - 1/4t^2 & 5/4 \leqq t. \end{cases}$$

Let $\omega_n^*(t)$ be the least upper bound of $P(\bar{X}_n \geqq t\mu)$ when $X_1, \cdots, X_n$ are independent and nonnegative with common mean μ. Clearly, $\omega_n(t) \leqq \omega_n^*(t)$. From [1] (in particular, Corollary 2.2) we have

$$\omega_n^*(t) \leqq \begin{cases} \dfrac{1}{t} - \dfrac{1}{4t^2} & \dfrac{3 + \sqrt{5}}{4} \leqq t; \quad n \text{ even}; \\[2ex] \dfrac{1}{t} - \dfrac{n^2 - 1}{n^2}\dfrac{1}{4t^2} & \dfrac{3n + 1 + (5n^2 + 6n + 5)^{1/2}}{4n} \leqq t; \quad n \text{ arbitrary}. \end{cases}$$

On the other hand, for any random variables X_i which satisfy our assumptions, $P(\bar{X}_n \geqq t\mu)$ is a lower bound for $\omega_n(t)$. In particular, if $nt \geqq 1$ and $X_i = 0$ or nt with respective probabilities $1 - 1/nt$ and $1/nt$, we get

$$\omega_n(t) \geqq 1 - \left(1 - \frac{1}{nt}\right)^n > \frac{1}{t} - \frac{n-t}{2n}\frac{1}{t^2}.$$

Hence we have for all positive integers n

$$\omega_n(t) = \frac{1}{t} - \frac{1+\theta}{4}\frac{n-1}{n}\frac{1}{t^2} \quad \text{if} \quad \frac{3n+1+(5n^2+6n+5)^{1/2}}{4n} \leqq t, \quad \frac{1}{n} \leqq \theta < 1; \tag{4.1}$$

$$\omega_{2n}(t) = \frac{1}{t} - \frac{1+\theta'}{4}\frac{1}{t^2} \quad \text{if} \quad \frac{5}{4} \leqq t, \qquad 0 \leqq \theta' < 1 - \frac{2}{n}. \tag{4.2}$$

Equation (4.1) is also true for $\omega_n^*(t)$, and (4.2) holds for $\omega_{2n}^*(t)$ if $\frac{1}{4}(3 + \sqrt{5}) \leqq t$. Thus for large values of t the known bounds for $\omega_n(t)$ and $\omega_n^*(t)$ cannot be improved substantially.

REFERENCES

[1] Z. W. Birnbaum, J. Raymond, and H. S. Zuckerman, "A generalization of Tchebychev's inequality to two dimensions", *Ann. Math. Stat.*, Vol. 18 (1947), pp. 70–79.

[2] E. J. Gumbel, "The maxima of the mean largest value and of the range", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 76–84.

[3] H. O. Hartley and H. A. David, "Universal bounds for mean range and extreme observation", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 85–99.

[4] W. Hoeffding, "The extrema of the expected value of a function of independent random variables", *Ann. Math. Stat.*, Vol. 26 (1955), pp. 268–275.

[5] R. L. Plackett, "Limits of the ratio of mean range to standard deviation", *Biometrika*, Vol. 34 (1947), pp. 120–122.

[6] H. Robbins, "Some remarks on the inequality of Tchebycheff", *Courant Anniversary Volume*, 1948, pp. 345–350.

SOME MINIMAX INVARIANT PROCEDURES FOR ESTIMATING A CUMULATIVE DISTRIBUTION FUNCTION[1]

BY OM P. AGGARWAL

Purdue University and University of Washington

1. Summary. Some invariant procedures, which are essentially step-functions, are considered as estimators of the cumulative distribution function of a one-dimensional random variable on which a finite fixed number of observations are given, for various loss functions. Two principal classes of loss functions are considered and it is shown that for a special loss function in one class the optimum procedure is the usual sample cumulative function.

2. Introduction. Suppose that a sample $X_1, X_2, \cdots, X_n$ of a one-dimensional chance variable X is given. In a recent paper, Birnbaum [1] has discussed various techniques for deciding whether X has a completely specified continuous cumulative distribution function (c.d.f.), $H(x) = P(X \leqq x)$. In this paper is discussed an allied problem, viz., that if $F(x) = P(X \leqq x)$ is the unknown continuous c.d.f. of X and if $\hat{F}(x)$ be an estimate of $F(x)$ based on the sample $X_1, X_2, \cdots, X_n$, what would be the best estimate $\hat{F}$ when certain forms of the loss function are given.

Consider the loss function

$$L(F, \hat{F}) = \int_{-\infty}^{\infty} |F(x) - \hat{F}(x)|^r \, dx, \tag{1}$$

where r is an integer $\geqq 1$. It is almost obvious that the only invariant procedures for estimating F under the group of all one-to-one monotone transformations of the real numbers onto themselves which leave the sample values X_i $(i = 1, 2, \cdots, n)$ invariant are those which estimate $F(x)$ by a step-function

$$\hat{F}(x) = \text{constant}, \quad \text{say } c_j \text{ for } \quad X^{(j)} \leqq x < X^{(j+1)}, \tag{2}$$

where $X^{(1)} < X^{(2)} < \cdots < X^{(n)}$ are the ordered observations and $X^{(0)}$ and $X^{(n+1)}$ denote $-\infty$ and $+\infty$ respectively.

Using this estimate $\hat{F}$, we get

$$\begin{aligned} L(F, \hat{F}) &= \sum_{j=0}^{n} \int_{X^{(j)}}^{X^{(j+1)}} |F(x) - c_j|^r \, dF(x) \\ &= \frac{1}{r+1} \sum_{j=0}^{n} [(F(X^{(j+1)}) - c_j)|F(X^{(j+1)}) - c_j|^r \\ &\qquad - (F(X^{(j)}) - c_j)|F(X^{(j)}) - c_j|^r] \end{aligned} \tag{3}$$

Received August 20, 1954; revised February 28, 1955.

[1] Work done under the sponsorship of the Office of Naval Research.

and the right-hand side of this equation is a symmetric function of $F(X_1), F(X_2), \cdots, F(X_n)$ where $X_1, X_2, \cdots, X_n$ is the unordered sample. Using the probability integral transformation, it is clear that the distribution of $L(F, \hat{F})$ does not depend on F for F continuous. Hence the risk R, being the expectation of L with respect to the distribution F, is constant and independent of F itself. We can thus take F to be a rectangular distribution over (0, 1) and write

$$R = E \sum_{j=0}^{n} \int_{X_j}^{X_{j+1}} |x - c_j|^r \, dx, \tag{4}$$

where $X_1 < X_2 < \cdots < X_n$ is an ordered sample of size n from this rectangular distribution over (0, 1), X_0 and X_{n+1} denote 0 and 1 respectively, and the symbol E denotes that the expectation is taken with respect to the rectangular distribution over (0, 1). In the rest of this paper, we shall use consistently the letter E to denote the fact that the expectation is to be taken with respect to the rectangular distribution over (0, 1).

The same argument applies when the loss function is of the form

$$L(F, \hat{F}) = \int_{-\infty}^{\infty} \frac{|F(x) - \hat{F}(x)|^r}{F(x)[1 - F(x)]} \, dF(x) \tag{5}$$

and in this case by taking $\hat{F}$ as in (2) we obtain

$$R = E \sum_{j=0}^{n} \int_{X_j}^{X_{j+1}} \frac{|x - c_j|^r}{x(1 - x)} \, dx \tag{6}$$

where X_j, $j = 0, 1, \cdots n + 1$, are the same as in (4).

It is obvious that since risk R is constant, a minimax procedure among the class of invariant procedures being considered will be to choose c_j, $j = 0, 1, \cdots, n$, such that R is minimum. We consider in this paper the values of c_j when the loss function is of the form (1) for all integers $r \geqq 1$ and when the loss function is of the form (5) for $r = 1$ and when r is an even integer $\geqq 2$. The case when r is odd in (5) seems to be rather complicated.

3. The loss function $L(F, \hat{F}) = \int_{-\infty}^{\infty} [F(x) - \hat{F}(x)]^r \, dF(x)$ where r is any positive even integer. Let $r = 2s$, then

$$R = E \sum_{j=0}^{n} \int_{X_j}^{X_{j+1}} (x - c_j)^{2s} \, dx = \sum_{j=0}^{n} Q_j, \tag{7}$$

where

$$Q_j = \frac{1}{2s + 1} E \sum_{k=0}^{2s+1} \binom{2s + 1}{k} (-c_j)^{2s+1-k} (X_{j+1}^k - X_j^k) \tag{8}$$

for $j = 0, 1, 2, \cdots, n$.

Since the distribution of the jth order statistic X_j in a sample of size n from the rectangular distribution over (0, 1) is a beta distribution with probability density

$$p(y) = \frac{1}{B(j, n - j + 1)} y^{j-1}(1 - y)^{n-j}, \qquad 0 \leqq y \leqq 1, \tag{9}$$

it is easily seen that for any positive integer r,

$$E(X_j^r) = \frac{j(j+1)\cdots(j+r-1)}{(n+1)(n+2)\cdots(n+r)}, \tag{10}$$

$$E(X_{j+1}^r - X_j^r) = \begin{cases} \dfrac{r(j+1)(j+2)\cdots(j+r-1)}{(n+1)(n+2)\cdots(n+r)} & \text{for } r \neq 1, \\[2ex] \dfrac{1}{n+1} \quad \text{for } r = 1. & j = 0, 1, \cdots, n. \end{cases} \tag{11}$$

Substituting from (11) in (8) we obtain

$$\begin{aligned} Q_j = \frac{1}{n+1} c_j^{2s} + \frac{1}{2s+1} \sum_{k=2}^{2s+1} \binom{2s+1}{k} \\ \cdot (-c_j)^{2s+1-k} \frac{k(j+1)\cdots(j+k-1)}{(n+1)\cdots(n+k)} \\ = \frac{1}{n+1}\left[c_j^{2s} + \sum_{k=2}^{2s+1} \binom{2s}{k-1} (-c_j)^{2s+1-k} \frac{(j+1)\cdots(j+k-1)}{(n+2)\cdots(n+k)} \right]. \end{aligned} \tag{12}$$

For conciseness we introduce the following notation somewhat similar to the binomial and distinguished from it by an asterisk. Let

$$\left(t - \frac{a+1}{b+1}\right)^{q*} = t^q + \sum_{k=1}^{q} (-1)^k \binom{q}{k} t^{q-k} \prod_{i=1}^{k} \frac{a+i}{b+i}, \tag{13}$$

for fixed real a and b and a positive integer q. For $q = 0$, let (13) be equal to 1. It is easily verified that for any positive integer r,

$$\frac{d^r}{dt^r}\left(t - \frac{a+1}{b+1}\right)^{q*} = \begin{cases} q(q-1)\cdots(q-r+1) \\ \quad \cdot \left(t - \dfrac{a+1}{b+1}\right)^{(q-r)*} & \text{when } r \leqq q, \\ 0 \quad \text{when } r > q. \end{cases} \tag{14}$$

Using this notation we can write

$$Q_j = \frac{1}{n+1}\left(c_j - \frac{j+1}{n+2}\right)^{2s*}. \tag{15}$$

We have to choose c_j so as to minimize R. Since $R = \sum_0^n Q_j$, and from (7) we see that for each j, Q_j is positive and depends only on j, it is obvious that minimizing R is equivalent to minimizing Q_j separately for each j. We obtain

$$\frac{\partial Q_j}{\partial c_j} = \frac{2s}{n+1}\left(c_j - \frac{j+1}{n+2}\right)^{(2s-1)*}, \tag{16}$$

$$\frac{\partial^2 Q_j}{\partial c_j^2} = \frac{2s(2s-1)}{n+1}\left(c_j - \frac{j+1}{n+2}\right)^{(2s-2)*}. \tag{17}$$

Since $Q_j = E \int_{X_j}^{X_{j+1}} (x - c_j)^{2s} dx > 0$, it is clear that

$$\frac{\partial^2 Q_j}{\partial c_j^2} = 2s(2s - 1)E \int_{X_j}^{X_{j+1}} (x - c_j)^{2s-2} dx > 0. \tag{18}$$

Let $f(c_j) = \partial Q_j / \partial c_j$. It is easily seen that $f(0)$ is negative and $f(1)$ is positive, and since $f'(c_j) > 0$ for all real c_j, $f(c_j)$ is a strictly increasing function of c_j. Hence $f(c_j) = 0$ for one and only one real value of c_j, and this c_j necessarily lies between 0 and 1. Thus we find that Q_j, and hence R, is minimized by setting $\partial Q_j / \partial c_j = 0$ and solving for c_j the resulting equation

$$\left(c_j - \frac{j+1}{n+2}\right)^{(r-1)*} = 0. \tag{19}$$

This equation has one and only one real root which lies between 0 and 1. The minimax invariant procedure for the loss function of this section is thus to estimate $F(x)$ by

$$\hat{F}(x) = c_j; \quad X_j \leqq x < X_{j+1}, \qquad j = 0, 1, \cdots, n, \tag{20}$$

where X_j, $j = 0, 1, \cdots, n + 1$, have been defined earlier and c_j is the real root of (19). It can further be seen from (19) that the equation remains unchanged if we replace j by $n - j$ and c_j by $1 - c_j$. Hence $c_{n-j} = 1 - c_j$, and we see that in practice the number of equations to be solved is about half the sample size.

Special case for $r = 2$. When $r = 2$, the equation (19) reduces to a linear equation

$$\left(c_j - \frac{j+1}{n+2}\right)^{1*} = 0, \tag{21}$$

which has the unique solution $c_j = (j + 1) / (n + 2)$. This result can, however, be obtained directly by writing the risk R from (7) and (12) for $s = 1$ in the form

$$R = \frac{1}{6(n+2)} + \frac{1}{n+1} \sum_{j=0}^{n} \left(c_j - \frac{j+1}{n+2}\right)^2. \tag{22}$$

We see thus that R is minimized by choosing

$$c_j = \frac{j+1}{n+2}, \qquad j = 0, 1, \cdots, n, \tag{23}$$

and hence the minimax invariant procedure is to estimate $F(x)$ by

$$\hat{F}(x) = \frac{j+1}{n+2}, \quad X_j \leqq x < X_{j+1}, \qquad j = 0, 1, \cdots, n, \tag{24}$$

where $(X_1, X_2, \cdots, X_n)$ is the ordered sample and X_0 and X_{n+1} stand for $-\infty$ and $+\infty$ respectively.

The minimum risk corresponding to this procedure is seen to be $\frac{1}{6}(n + 2)$. It is of some interest to note that the risk corresponding to the usual procedure of taking $c_j = j/n$ is given by $\frac{1}{6}n$.

4. The loss function $L(F, \hat{F}) = \int_{-\infty}^{\infty} |F(x) - \hat{F}(x)|^r \, dF(x)$, where r is any positive integer. In this case

$$R = E \sum_{j=0}^{n} \int_{X_j}^{X_{j+1}} |x - c_j|^r \, dx = \sum_{j=0}^{n} Q_j, \tag{25}$$

where

$$Q_j = \frac{1}{r+1} E[(X_{j+1} - c_j)\,|X_{j+1} - c_j|^r - (X_j - c_j)\,|X_j - c_j|^r]. \tag{26}$$

Using (9) we obtain

$$\begin{aligned} E[(X_j - c_j)\,|X_j - c_j|^r] = j\binom{n}{j}\Big[\int_{c_j}^{1} (y - c_j)^{r+1} y^{j-1}(1 - y)^{n-j} \, dy \\ - \int_0^{c_j} (c_j - y)^{r+1} y^{j-1}(1 - y)^{n-j} \, dy\Big], \end{aligned} \tag{27}$$

and similarly,

$$\begin{aligned} E[(X_{j+1} - c_j)\,|X_{j+1} - c_j|^r] = (n - j)\binom{n}{j} \\ \cdot \Big[\int_{c_j}^{1} (y - c_j)^{r+1} y^{j}(1 - y)^{n-j-1} \, dy - \int_0^{c_j} (c_j - y)^{r+1} y^{j}(1 - y)^{n-j-1} \, dy\Big]. \end{aligned} \tag{28}$$

From (27) and (28) we obtain

$$\begin{aligned} Q_j = \frac{1}{r+1}\binom{n}{j}\Big[\int_{c_j}^{1} (y - c_j)^{r+1} y^{j-1}(1 - y)^{n-j-1}(ny - j) \, dy \\ + (-1)^r \int_0^{c_j} (y - c_j)^{r+1} y^{j-1}(1 - y)^{n-j-1}(ny - j) \, dy\Big]. \end{aligned} \tag{29}$$

Again it is obvious that to minimize R is equivalent to minimizing Q_j for each j. Further we see that the conditions for differentiation with respect to c_j under the integral sign in (29) are satisfied, and we obtain

$$\begin{aligned} \frac{\partial Q_j}{\partial c_j} = -\binom{n}{j}\Big[\int_{c_j}^{1} (y - c_j)^{r} y^{j-1}(1 - y)^{n-j-1}(ny - j) \, dy \\ + (-1)^r \int_0^{c_j} (y - c_j)^{r} y^{j-1}(1 - y)^{n-j-1}(ny - j) \, dy\Big], \end{aligned} \tag{30}$$

$$\begin{aligned} \frac{\partial^2 Q_j}{\partial c_j^2} &= r\binom{n}{j}\Big[\int_{c_j}^{1} (y - c_j)^{r-1} y^{j-1}(1 - y)^{n-j-1}(ny - j) \, dy \\ &\qquad + (-1)^r \int_0^{c_j} (y - c_j)^{r-1} y^{j-1}(1 - y)^{n-j-1}(ny - j) \, dy\Big] \\ &= \begin{cases} r(r-1)E \displaystyle\int_{X_j}^{X_{j+1}} |x - c_j|^{r-2} \, dx, & \text{for } r \geqq 2 \\ 2\dbinom{n}{j} c_j^j(1 - c_j)^{n-j}, & \text{for } r = 1. \end{cases} \end{aligned} \tag{31}$$

Define a function f by $f(c_j) = \partial Q_j / \partial c_j$. We see by straightforward computations that

$$f(0) = -r \frac{n!(r+j-1)!}{j!(r+n)!} < 0,$$

$$f(1) = r \frac{n!(r+n-j-1)!}{(n-j)!(r+n)!} > 0.$$

Since from (31) it is seen that, for all $r \geqq 2$, $f'(c_j) = \partial^2 Q_j / \partial c_j^2 > 0$ for all real c_j (the special case for $r = 1$ is given at the end of this section), f is a strictly increasing function of c_j and assumes the value zero for one and only one real value of c_j, and this value of c_j necessarily lies between zero and one. Thus we find that Q_j and hence R is minimized by setting $\partial Q_j / \partial c_j = 0$ and solving for c_j the resulting equation

$$\begin{aligned} &\int_{c_j}^1 (y - c_j)^r y^{j-1}(1-y)^{n-j-1}(ny - j)\, dy \\ (32) \qquad &\qquad + (-1)^r \int_0^{c_j} (y - c_j)^r y^{j-1}(1-y)^{n-j-1}(ny - j)\, dy = 0. \end{aligned}$$

Thus the problem reduces to that of solving the above equation for $j = 0, 1, \cdots, n$. The general solution of (32) giving c_j explicitly in terms of j, n, and r does not seem to be possible. We shall, however, simplify the equation so that it should not be too difficult to obtain the solution in any given case. It can, however, be proved from (32) that $c_{n-j} = 1 - c_j$, so that the number of equations to be solved in practice will be about half the sample size.

We can write (32) as

$$\begin{aligned} &\int_0^1 (y - c_j)^r y^{j-1}(1-y)^{n-j-1}(ny - j)\, dy \\ (33) \qquad &\qquad = [1 - (-1)^r] \int_0^{c_j} (y - c_j)^r y^{j-1}(1-y)^{n-j-1}(ny - j)\, dy. \end{aligned}$$

The left-hand side of equation (33) can be expressed as

$$(34) \qquad \sum_{k=0}^{r} k \binom{r}{k} (-c_j)^{r-k} B(j + k, n - j + 1),$$

which indicates that the coefficient of c_j^r is zero. For $k \neq 0$, we can utilize the fact that $\binom{r}{k} k = r \binom{r-1}{k-1}$ and reduce it further to the form

$$(35) \quad rB(j, n - j + 1) \sum_{k=1}^{r} \binom{r-1}{k-1} (-c_j)^{r-k} \frac{j(j+1) \cdots (j+k-1)}{(n+1)(n+2) \cdots (n+k)},$$

which by making use of the notation introduced in (13) can be written as

$$(36) \qquad (-1)^{r-1} rB(j + 1, n - j + 1) \left(c_j - \frac{j+1}{n+2}\right)^{(r-1)*}.$$

When r is even, the right-hand side of the equation (33) reduces to zero and cancelling out the nonzero coefficient $(-1)^{r-1}rB(j+1, n-j+1)$ from the left-hand side as expressed by (36) we obtain c_j as a root of the same equation as (19) obtained earlier by a different method.

The right-hand side of the equation (33), except for the factor $[1-(-1)^r]$, can be written as

$$(37) \qquad \sum_{k=0}^{r} \binom{r}{k} (-c_j)^{r-k} \int_0^{c_j} \sum_{s=0}^{n-j} (-1)^{s-1}(j+s) \binom{n-j}{s} y^{k+j+s-1}\, dy,$$

and by making use of the relation

$$(38) \qquad \sum_{k=0}^{r} (-1)^k \binom{r}{k} \frac{1}{k+t} = B(t, r+1),$$

it can be reduced to

$$(39) \qquad (-1)^{r-1} r \sum_{s=0}^{n-j} (-1)^s \binom{n-j}{s} B(r, j+s+1) c_j^{r+j+s}.$$

Using (36) and (37) we can, thus, write the equation (33) as

$$(40) \qquad \begin{aligned} &B(j+1, n-j+1)\left(c_j - \frac{j+1}{n+2}\right)^{(r-1)*} \\ &\qquad = [1-(-1)^r] \sum_{s=0}^{n-j} (-1)^s \binom{n-j}{s} B(r, j+s+1) c_j^{r+j+s}. \end{aligned}$$

This equation is to be solved for c_j to get a minimax invariant procedure for estimating F when the loss function is given by (1). When r is even, the factor $1-(-1)^r = 0$ and we get an equation of degree $(r-1)$. When r is odd, the factor $1-(-1)^r = 2$ and the equation reduces to

$$(41) \qquad \begin{aligned} &\sum_{s=0}^{n-j} (-1)^s \binom{n-j}{s} B(r, j+s+1) c_j^{r+j+s} \\ &\qquad - \tfrac{1}{2} B(j+1, n-j+1)\left(c_j - \frac{j+1}{n+2}\right)^{(r-1)*} = 0 \end{aligned}$$

which is an equation of degree $n+r$. In either case there is one and only one real root which lies between 0 and 1 and the set of such roots for $j = 0, 1, \cdots, n$ minimizes R.

An alternative way of expressing the right-hand side of (33) is to rewrite (39) in the form:

$$(42) \qquad (-1)^{r-1} r! \sum_{s=0}^{n-j} (-1)^s \binom{n-j}{s} \frac{c_j^{r+j+s}}{(j+s+1)(j+s+2)\cdots(j+s+r)}.$$

It is easily verified that (42) is equal to

$$(43) \qquad \begin{aligned} &(-1)^{r-1} r! \sum_{s=0}^{n-j} (-1)^s \binom{n-j}{s} \int_0^{c_j} \int_0^{z_r} \cdots \int_0^{z_2} z_1^{j+s}\, dz_1 \cdots dz_r \\ &\qquad = (-1)^{r-1} r! \int_0^{c_j} \int_0^{z_r} \cdots \int_0^{z_2} z_1^j (1-z_1)^{n-j}\, dz_1 \cdots dz_r. \end{aligned}$$

The equation (40) can, therefore, also be expressed as

$$(44)\quad \begin{aligned} &B(j+1, n-j+1)\left(c_j - \frac{j+1}{n+2}\right)^{(r-1)*} \\ &= [1-(-1)^r](r-1)!\int_0^{c_j}\int_0^{z_r}\cdots\int_0^{z_2} z_1^j(1-z_1)^{n-j}\,dz_1\cdots dz_r. \end{aligned}$$

Special case for $r = 1$. When $r = 1$, (30) is easily seen to reduce to

$$(45)\qquad \frac{\partial Q_j}{\partial c_j} = 2\binom{n}{j}\left[\int_0^{c_j} z^j(1-z)^{n-j}\,dz - \tfrac{1}{2}B(j+1, n-j+1)\right],$$

from which follows easily the result given in (31), viz.,

$$(46)\qquad \frac{\partial^2 Q_j}{\partial c_j^2} = 2\binom{n}{j} c_j^j(1-c_j)^{n-j}.$$

Setting $\partial Q_j / \partial c_j = 0$ and solving we obtain c_j as the median of the beta distribution with density

$$(47)\qquad g(z) = \frac{1}{B(j+1, n-j+1)} z^j(1-z)^{n-j}, \qquad 0 \leqq z \leqq 1,$$

for $j = 0, 1, 2, \cdots, n$. Since (46) shows that $\partial^2 Q_j / \partial c_j^2 > 0$ for $0 < c_j < 1$, it follows that this solution for c_j in fact minimizes Q_j for $j = 0, 1, \cdots, n$, and hence minimizes R. The equation (44) for c_j obtained for $r \geqq 2$ thus holds good for $r = 1$ as well and the minimax invariant procedure is seen to estimate $F(x)$ by

$$(48)\qquad \hat{F}(x) = c_j; \qquad X_j \leqq x < X_{j+1}, \qquad j = 0, 1, \cdots, n,$$

where $(X_1, X_2, \cdots, X_n)$ is the ordered sample, X_0 and X_{n+1} stand for $-\infty$ and $+\infty$ respectively, and c_j $(j = 0, 1, \cdots, n)$ is the median of the beta distribution with density (47). It is rather interesting to note that the value $(j+1)/(n+2)$ for c_j obtained in the last section for $r = 2$ is the mean of the same beta distribution.

The actual computation of the values of c_j $(j = 0, 1, \cdots, n)$ can be easily carried out, for a given n, with the help of the tables of the incomplete beta function [2]. In the notation of those tables

$$(49)\qquad I_x(p, q) = \frac{\int_0^x x^{p-1}(1-x)^{q-1}\,dx}{\int_0^1 x^{p-1}(1-x)^{q-1}\,dx}.$$

Thus we have to find the value of x for each j such that

$$(50)\qquad I_x(j+1, n-j+1) = \tfrac{1}{2}.$$

Using the relation

$$(51)\qquad I_x(p, q) = 1 - I_{1-x}(q, p),$$

TABLE I

Values of c_j $(j = 0, 1, \cdots, n)$ for $n = 1, 2, \cdots, 12$

n	c_j												
	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
1	.29	.71											
2	.21	.50	.79										
3	.16	.39	.61	.84									
4	.13	.31	.50	.69	.87								
5	.11	.26	.42	.58	.74	.89							
6	.09	.23	.36	.50	.64	.77	.91						
7	.08	.20	.32	.44	.56	.68	.80	.92					
8	.07	.18	.28	.39	.50	.61	.72	.82	.93				
9	.07	.16	.26	.35	.45	.55	.65	.74	.84	.93			
10	.06	.15	.23	.32	.41	.50	.59	.68	.77	.85	.94		
11	.06	.14	.22	.30	.38	.46	.54	.62	.70	.78	.86	.94	
12	.05	.13	.20	.27	.35	.42	.50	.58	.65	.73	.80	.87	.95

it is seen that as in the general case,

$$c_{n-j} = 1 - c_j. \tag{52}$$

The values of c_j $(j = 0, 1, \cdots, n)$ for $n = 1, 2, \cdots, 12$ correct to two decimal places are computed and tabulated as shown in Table I.

5. The loss function $L(F, \hat{F}) = \int_{-\infty}^{\infty} [F(x) - \hat{F}(x)]^r / F(x)[1 - F(x)]\, dF(x)$ where r is any positive even integer.

Let $r = 2s$; then

$$R = E \sum_{j=0}^{n} \int_{X_j}^{X_{j+1}} \frac{(x - c_j)^{2s}}{x(1 - x)}\, dx = \sum_{j=0}^{n} Q_j, \tag{53}$$

where

$$Q_j = E \int_{X_j}^{X_{j+1}} \frac{(x - c_j)^{2s}}{x(1 - x)}\, dx. \tag{54}$$

Since $X_0 = 0$ and $X_{n+1} = 1$, it is clear that in order to obtain finite risk it is necessary and sufficient that $c_0 = 0$ and $c_n = 1$. For $j \neq 0, n$, we can write

$$\begin{aligned} Q_j = E\Bigg[\sum_{h=0}^{2s-2} \frac{1}{h+1} a_h (X_{j+1}^{h+1} - X_j^{h+1}) &+ c_j^{2s}(\log X_{j+1} - \log X_j) \\ &- (1 - c_j)^{2s}\{\log (1 - X_{j+1}) - \log (1 - X_j)\}\Bigg], \end{aligned} \tag{55}$$

where

$$a_h = - \sum_{i=0}^{2s-2-h} \binom{2s}{i} (-c_j)^i; \qquad h = 0, 1, 2, \cdots, 2s - 2. \tag{56}$$

The probability density of X_j is given by (9), from which we obtain

$$E(\log X_j) = j\binom{n}{j}\int_0^1 y^{j-1}(1-y)^{n-j}\log y\,dy. \tag{57}$$

In order to evaluate (57) we use the following lemma.

LEMMA 5.1.

$$\int_0^1 y^{j-1}(1-y)^{n-j}\log y\,dy = \frac{\Gamma(j)\Gamma(n-j+1)}{\Gamma(n+1)}[\psi(j)-\psi(n+1)], \tag{58}$$

where $\psi(k) = \Gamma'(k)/\Gamma(k)$.

PROOF. Let $f(\alpha) = \int_0^1 y^{\alpha-1}(1-y)^{n-j}\,dy$. The left-hand side of (58) is $f'(\alpha)$ evaluated at $\alpha = j$ as can be seen by differentiating under the integral sign. But $f(\alpha) = \Gamma(\alpha)\Gamma(n-j+1)/\Gamma(\alpha+n-j+1)$, and the desired result is obtained by evaluating the logarithmic derivative of $f(\alpha)$ at $\alpha = j$.

From the lemma 5.1 and (57) we get

$$E(\log X_j) = \psi(j) - \psi(n+1). \tag{59}$$

In the same way, we obtain

$$E\log(1-X_j) = \psi(n-j+1) - \psi(n+1). \tag{60}$$

Further, since $\Gamma(k+1) = k\Gamma(k)$, $\Gamma'(k+1) = \Gamma(k) + k\Gamma'(k)$, we see that $\psi(k+1) = \Gamma'(k+1)/\Gamma(k+1) = 1/k + \psi(k)$, and hence the function ψ satisfies the difference equation

$$\psi(k+1) - \psi(k) = 1/k. \tag{61}$$

From (59), (60), and (61) we get

$$E(\log X_{j+1} - \log X_j) = 1/j, \qquad \text{for } j = 1, 2, \cdots, n, \tag{62}$$

and

$$E[\log(1-X_{j+1}) - \log(1-X_j)] = -1/(n-j), \qquad \text{for } j = 0, 1, \cdots, n-1. \tag{63}$$

Substituting from (11), (62), and (63) in (55) we get

$$Q_j = \sum_{h=0}^{2s-2}\frac{(j+h)!\,n!}{(n+h+1)!\,j!}a_h + \frac{1}{j}c_j^{2s} + \frac{1}{n-j}(1-c_j)^{2s}, \tag{64}$$

and substituting from (56), we can write

$$Q_j = \frac{n!}{j!}\sum_{h=0}^{2s-2}\frac{(j+h)!}{(n+h+1)!}\sum_{i=0}^{2s-2-h}(-1)^{i+1}\binom{2s}{i}c_j^i + \frac{1}{j}c_j^{2s} + \frac{1}{n-j}(1-c_j)^{2s}. \tag{65}$$

This is a $2s$th degree polynomial in c_j. Collecting the coefficients of like powers of c_j we obtain, for $k = 0, 1, 2, \cdots, 2s-2$,

$$Q_j = \frac{n}{j(n-j)}c_j^{2s} - \frac{2s}{n-j}c_j^{2s-1} + \sum_{k=0}^{2s-2} g_k c_j^k, \tag{66}$$

where

$$g_k = (-1)^{k+1}\binom{2s}{k}\left[\frac{n!}{j!}\sum_{h=0}^{2s-2-k}\frac{(j+h)!}{(n+h+1)!} - \frac{1}{n-j}\right]. \tag{67}$$

To simplify (66) further, we state and prove the following lemma.

LEMMA 5.2. *If j and n are positive integers and $j < n$, then*

$$\frac{n!}{j!}\sum_{h=0}^{q}\frac{(j+h)!}{(n+h+1)!} = \frac{1}{n-j}\left[1 - \prod_{\alpha=1}^{q+1}\frac{j+\alpha}{n+\alpha}\right]. \tag{68}$$

PROOF. The left-hand side is equal to

$$\binom{n}{j}\sum_{h=0}^{q}\frac{(j+h)!\,(n-j)!}{(n+h+1)!} = \binom{n}{j}\sum_{h=0}^{q}\int_0^1 x^{j+h}(1-x)^{n-j}\,dx$$

$$= \binom{n}{j}\int_0^1 (x^j - x^{j+q+1})(1-x)^{n-j-1}\,dx$$

$$= \text{the right-hand side, after simplification.}$$

Substituting in (67) from (68) when $q = 2s - 2 - k$, we obtain

$$g_k = (-1)^k\frac{1}{n-j}\binom{2s}{k}\prod_{\alpha=1}^{2s-1-k}\frac{j+\alpha}{n+\alpha} \quad \text{for } k = 0, 1, 2, \cdots, 2s-2, \tag{69}$$

and substituting now in (66) we obtain

$$\begin{aligned} Q_j &= \frac{n}{j(n-j)}\left[c_j^{2s} + \sum_{k=0}^{2s-1}\binom{2s}{k}(-c_j)^k\prod_{\alpha=0}^{2s-1-k}\frac{j+\alpha}{n+\alpha}\right] \\ &= \frac{n}{j(n-j)}\left(c_j - \frac{j}{n}\right)^{2s*}, \end{aligned} \tag{70}$$

using the notation introduced in (13).

Now with the same reasoning as in Section 3 it will be seen that Q_j and hence R is minimized by setting $\partial Q_j / \partial c_j = 0$ and solving for c_j the resulting equation

$$(c_j - j/n)^{(r-1)*} = 0. \tag{71}$$

This equation, by the same argument as in Section 3, has one and only one real root which lies between 0 and 1. Since for $j = 0$, (71) reduces to $c_0^{r-1} = 0$ giving $c_0 = 0$ as the only real root, and for $j = n$, it reduces to $(c_n - 1)^{r-1} = 0$, giving $c_n = 1$ as the only real root, it follows that we can say that the minimax invariant procedure for the loss function of this section is to estimate $F(x)$ by

$$\hat{F}(x) = c_j\,; \qquad X_j \leq x < X_{j+1}\,, \qquad j = 0,\ 1, \cdots,\ n,$$

where X_j, $j = 0, 1, \cdots, n+1$, have been defined earlier and c_j is the real root of (71). Again the number of equations to be solved in practice will be about half the sample size since it can be easily seen that (71) remains unchanged by replacing j by $n - j$ and c_j by $1 - c_j$, so that $c_{n-j} = 1 - c_j$.

Special case for $r = 2$. When $r = 2$, the equation (71) reduces, for each j, to a linear equation

$$(c_j - j/n)^{1^*} = 0,$$

which has the unique solution $c_j = j/n$. This can also be seen by using (35), (70), (62), and (63) for $r = 2$ and writing the risk R in the form

$$R = \frac{1}{n} + \sum_{j=1}^{n-1} \frac{n}{j(n-j)} \left(c_j - \frac{j}{n}\right)^2. \tag{72}$$

Thus the minimax invariant estimate $\hat{F}$ for the loss function in the special case for $r = 1$ in this section turns out to be the usual sample cumulative function

$$\hat{F}(x) = c_j = j/n, \quad \text{when} \quad X_j \leqq x < X_{j+1}, \qquad j = 0, 1, \cdots, n, \tag{73}$$

where $X_1 < X_2 < \cdots < X_n$ is an ordered sample from the c.d.f. F, X_0 and X_{n+1} standing for $-\infty$ and $+\infty$ respectively. The actual value of the risk corresponding to this estimate is $1/n$.

6. The loss function $L(F, \hat{F}) = \int_{-\infty}^{\infty} |F(x) - \hat{F}(x)| / F(x)[1 - F(x)]\, dF(x)$. In this case we obtain

$$R = E \sum_{j=0}^{n} \int_{X_j}^{X_{j+1}} |x - c_j| / x(1 - x)\, dx = \sum_{j=0}^{n} Q_j, \tag{74}$$

where

$$Q_j = E \int_{X_j}^{X_{j+1}} |x - c_j| / x(1 - x)\, dx. \tag{75}$$

As in the last section, it will be seen that for finite risk the necessary and sufficient condition is that $c_0 = 0$ and $c_n = 1$. For $j \neq 0, n$, we obtain

$$\begin{aligned} Q_j = E[c_j\, |\log c_j - \log X_j| - c_j\, |\log c_j - \log X_{j+1}| \\ + (1 - c_j)|\log(1 - c_j) - \log(1 - X_{j+1})| \\ - (1 - c_j)|\log(1 - c_j) - \log(1 - X_j)|]. \end{aligned} \tag{76}$$

The distribution of X_j has probability density $p(y)$ given by (9) and the distribution of X_{j+1} has the probability density

$$q(y) = \frac{1}{B(j + 1, n - j)} y^j (1 - y)^{n-j-1}, \qquad 0 \leqq y \leqq 1. \tag{77}$$

Using (9) and (77) we can express Q_j in the form

$$Q_j = \binom{n}{j} \left[\int_0^{c_j} g(c_j, y)\, dy - \int_{c_j}^{1} g(c_j, y)\, dy\right], \tag{78}$$

where

$$g(c_j, y) = [c_j \log c_j + (1 - c_j) \log (1 - c_j) - c_j \log y - (1 - c_j) \log (1 - y)]y^{j-1}(1 - y)^{n-j-1}(j - ny). \tag{79}$$

Straightforward integration leads to

$$\int g(c_j, y)\, dy = y^j(1 - y)^{n-j}[c_j(\log c_j - \log y) + (1 - c_j)(\log (1 - c_j) - \log (1 - y))] + \int (c_j - y)y^{j-1}(1 - y)^{n-j-1}\, dy + \text{constant}, \tag{80}$$

which enables us to obtain Q_j as

$$Q_j = \binom{n}{j}\left[\int_0^{c_j} (c_j - y)y^{j-1}(1 - y)^{n-j-1}\, dy - \int_{c_j}^1 (c_j - y)y^{j-1}(1 - y)^{n-j-1}\, dy\right], \tag{81}$$

for $j = 1, 2, \cdots, n - 1$. Since Q_0 and Q_n are fixed, and each Q_j is positive and depends only on j, minimizing R is equivalent to minimizing Q_j for each j. We see that

$$\frac{\partial Q_j}{\partial c_j} = \binom{n}{j}\int_0^{c_j} y^{j-1}(1 - y)^{n-j-1}\, dy - \binom{n}{j}\int_{c_j}^1 y^{j-1}(1 - y)^{n-j-1}\, dy, \tag{82}$$

$$\frac{\partial^2 Q_j}{\partial c_j^2} = 2\binom{n}{j}c_j^{j-1}(1 - c_j)^{n-j-1}. \tag{83}$$

Setting $\partial Q_j / \partial c_j = 0$ and solving we obtain c_j as the median of the beta distribution with density

$$h(z) = \frac{1}{B(j, n - j)}\, z^{j-1}(1 - z)^{n-j-1}, \qquad 0 \leqq z \leqq 1, \tag{84}$$

for $j = 1, 2, \cdots, n - 1$. Since (83) shows that $\partial^2 Q_j / \partial c_j^2 > 0$ for $0 < c_j < 1$, it follows that this solution for c_j in fact minimizes Q_j and hence minimizes R. To summarize, the minimax invariant procedure for the loss function considered in this section is to estimate $F(x)$ by

$$\hat{F}(x) = c_j\,; \qquad X_j \leqq x < X_{j+1}, \qquad j = 0, 1, \cdots, n, \tag{85}$$

where X_j, $j = 0, 1, \cdots, n + 1$, have been defined earlier, $c_0 = 0$, $c_n = 1$ and for $j = 1, 2, \cdots, n - 1$, c_j is the median of the beta distribution with density (84). Again it is interesting to note that the value j/n for c_j obtained in the last section for $r = 2$ is the mean of the same beta distribution.

Further it is obvious that $c_{n-j} = 1 - c_j$ and only about half the total number of c values are to be actually computed. These can be obtained with the help

of the tables of the incomplete beta-function [2] as indicated in Section 4. However, if a table for c values like Table I has been constructed, no fresh computations are needed, since the value of $c_j \quad (j = 1, 2, \cdots, n - 1)$ for any n in this case is equal to the value of c_{j-1} for $n - 2$ in Table I. For example, when $n = 10$, the values of $c_j \quad (j = 0, 1, \cdots, 10)$ correct to two decimal places are

$$0, .07, .18, .28, .39, .50, .61, .72, .82, .93, 1. \tag{86}$$

I am thankful to Professors Z. W. Birnbaum and H. Rubin for some helpful discussions during the preparation of this paper.

REFERENCES

[1] Z. W. Birnbaum, "Distribution-free tests of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 1–8.

[2] Karl Pearson, *Tables of the Incomplete Beta-Function*, The Biometrika Office, University College, London.

DISTRIBUTION OF QUADRATIC FORMS AND SOME APPLICATIONS[1]

By Arthur Grad and Herbert Solomon

Office of Naval Research, Washington, D. C.
and
Teachers College, Columbia University

1. Summary. The authors were prompted by a general problem concerning hit probabilities arising in military operations to seek the distribution of $Q_k = \sum_{i=1}^k a_i x_i^2$, $k = 2, 3$, where the x_i are normally and independently distributed with zero mean and unit variance, $\sum a_i = 1$, and $a_i > 0$. While the distribution of a positive definite quadratic form in independent normal variates has been the subject of several papers in recent years [6], [11], [12], laborious computations are required to prepare from existing results the percentiles of the distribution and a table of hit probabilities. This paper discusses the exact distribution of Q_k and then obtains and tabulates the distributions of Q_2 and Q_3, accurate to four places. Three other approaches to the distributions are discussed and compared with the exact results: a derivation by Hotelling [8], the Cornish-Fisher asymptotic approximation [3], and the approximation obtained by replacing the quadratic form with a chi-square variate whose first two moments are equated to those of the quadratic form—a type of approximation used in components of variance analysis. The exact values and the approximations are given in Tables I and II. The tables have been prepared with the original problem in mind, but also serve as an aid in several problems arising out of quite different contexts, [1], [2], [13]. These are discussed in Section 6.

2. Introduction. A general class of problems arises in military operations when the hit probability of a weapon depends on the combination of two random errors. Suppose random errors in predicted location or predicted position of target and random errors in aim of weapon occur. For purposes of exposition let us limit ourselves to errors in two dimensions. Denote the true position of a target by T, the predicted position, or point of aim, by A, and the point of impact of a weapon aimed at A by I. Let x_1, y_1, be the components of the vector TA and x_2, y_2 the components of the vector AI. If we denote the radius of effectiveness of the weapon by R, then the probability of a hit P is the probability that the resultant vector TI has length no greater than R, or

$$P = P\{x_3^2 + y_3^2 \leqq R^2\}, \tag{1}$$

where $x_3 = x_1 + x_2$, $y_3 = y_1 + y_2$.

Received July 2, 1954; revised April 18, 1955.

[1] The tables in this report were computed at Columbia University and Stanford University with the partial support of Office of Naval Research contracts N6onr 271 Task Order II (NR-042-034) and N6onr 251 Task Order III (NR-042-993).

TABLE I

$P(Q_2 \leqq t)$

t	α_2, α_1: .5, .5	.6, .4	.7, .3	.8, .2	.9, .1	.95, .05	.99, .01	1, 0
.1	09516 1384	09693	1029 1028 1285 1381	1158	1461 1345 2037 2368	1813	2359	2482
.2	1813 2023	1843	1943 1942 2126 2052	2153	2594 2465 2926 3114	3002	3384	3453
.3	2592 2691	2630	2757 2756 2871 2756	3011	3494 3399 3641 3811	3858	4115	4161
.4	3297 3345	3340	3482 3481 3542 3444	3755	4226 4180 4248 4436	4521	4697	4729
.5	3935 3963	3981	4128 4127 4146 4088	4402	4831 4835 4775 4986	5060	5182	5205
.6	4512 4533	4559	4705 4705 4693 4677	4968	5342 5387 5240 5465	5513	5599	5614
.7	5034 5052	5080	5221 5221 5187 5209	5464	5780 5854 5652 5883	5904	5962	5972
.8	5507 5523	5550	5682 5683 5633 5693	5901	6159 6251 6022 6249	6246	6283	6289
.9	5934 5950	5975	6095 6096 6037 6112	6287	6491 6592 6353 6572	6549	6570	6572

TABLE I—*Continued*

t	a_2, a_1							
	.5, .5	.6, .4	.7, .3	.8, .2	.9, .1	.95, .05	.99, .01	1, 0
1.0	6321	6358	6466	6630	6785	6819	68267	68269
			6467		6886			
			6402		6653			
	6336		6493		6859			
1.5	7769	7785	7826	7866	7858	7830	7801	7793
			7827		7900			
			7770		7781			
	7783		7881		7922			
2.0	8647	8646	8638	8604	8527	8478	8438	8427
			8638		8508			
			8606		8498			
	8749		8788		8700			
3.0	9502	9487	9441	9365	9269	9219	9178	9167
			9441		9234			
			9442		9283			
	9998		10000		9998			
4.0	9817	9802	9761	9698	9624	9585	9553	9545
			9760		9620			
			9770		9643			
	10000		10000		10000			
5.0	9933	9923	9895	9853	9803	9775	9753	9746
			9895		9812			
			9903		9817			
	10000		10000		10000			

First entry in cell is exact to 4 decimal places.
Second entry is Hotelling's result.
Third entry is "components of variance" chi square approximation.
Fourth entry is Cornish-Fisher result.

Now assume that the two random errors are each subject to a bivariate normal distribution with zero means and with covariance matrix $\|{}_p\sigma_{ij}\|$ and $\|{}_a\sigma_{ij}\|$ respectively. Then x_3 and y_3 are components of a vector having a bivariate normal distribution with zero means and covariance matrix $\|{}_p\sigma_{ij} + {}_a\sigma_{ij}\| = \|\lambda_{ij}\|$. For the present, assume the components of each error to be independent; i.e., $\|{}_p\sigma_{ij}\|$ and $\|{}_a\sigma_{ij}\|$ are diagonal. This restriction, which is not essential, implies that x_3 and y_3 are independently distributed. If $x = \lambda_{11}^{-1/2}\, x_3$ and $y = \lambda_{22}^{-1/2}\, y_3$, then x^2 and y^2 each have a chi-square distribution with one degree of freedom. We may then write

$$P = P\{a_1x^2 + a_2y^2 \leqq t\} \tag{2}$$

TABLE II

$P\{Q_3 \leqq t\}$

t	a_3, a_2, a_1								
	$\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$	.4,. 3, .3	.4, .4, .2	.5, .3, .2	.6, .2, .2	5, .4, .1	.6, .3, .1	.7, .2, .1	.8, .1, .1
.1	03997	04146	04313	04385	05035	05169	05421	06062	07419
		04048		04377				05564	05773
		0470		0602				1150	1548
		0697		0721				0945	1544
.2	10357	1053	1094	1123	1217	1282	1338	1477	1803
		1047		1122				1402	1483
		1083		1275				1971	2416
		1220		1265				1633	2336
.3	17457	1763	1830	1873	2026	2081	2162	2357	2758
		1763		1872				2296	2458
		1768		1985				2716	3155
		1849		1916				2411	3137
.4	24700	2491	2571	2624	2803	2852	2951	3179	3625
		2491		2623				3159	3406
		2474		2692				3397	3805
		2529		2617				3200	3886
.5	31773	3201	3287	3346	3541	3570	3679	3923	4353
		3201		3346				3952	4273
		3172		3375				4016	4381
		3216		3319				3946	4596
.6	38507	3875	3961	4023	4223	4228	4340	4584	4979
		3875		4024				4663	5037
		3841		4020				4580	4897
		3880		3992				4623	5141
.7		4505	4587	4649	4843	4825	4936	5169	5515
		4505		4650					
		4471		4621				4909	5360
		4506		4620				5214	5649
.8	50637	5086	5161	5220	5402	5363	5469	5683	5974
		5086		5222				5829	6239
		5056		5175				5555	5776
		5085		5195				5751	6088
.9		5618	5683	5739	5902	5848	5945	6136	6371
		5618		5740					
		5594		5682				5975	6152
		5615		5718				6175	6471

TABLE II—*Continued*

t	a_3, a_2, a_1								
	$\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$	.4, .3, .3	.4, .4, .2	.5, .3, .2	.6, .2, .2	.5, .4, .1	.6, .3, .1	.7, .2, .1	.8, .1, .1
1.0	60837	6102	6156	6206	6349	6282	6370	6535	6717
		6102		6207				6697	7056
		6083		6143				6355	6491
		6097		6180				6619	6806
1.5		7881	7884	7901	7935	7863	7895	7935	7930
		7881		7901					
		7885		7848				7776	7766
		7876		7894				8042	8008
2.0	88839	8879	8853	8844	8808	8770	8760	8723	8663
		8879		8844				8659	8527
		8889		8820				8636	8558
		8972		8931				8992	8888
3.0	97071	9698	9668	9645	9577	9591	9552	9477	9378
		9698		9645				9394	9270
		9702		9650				9477	9379
		10000		10000				9933	10000
4.0	99262	9920	9905	9888	9841	9863	9831	9775	9703
		9920		9888				9763	9734
		9921		9896				9794	9724
		10000		10000				10000	10000
5.0	99818	9979	9973	9963	9938	9954	9935	9900	9855
		9979		9964				9916	9897
		9979		9969				9917	9874
		10000		10000				10000	10000

First entry in cell is exact to 4 decimal places.
Second entry is Hotelling's result.
Third entry is "components of variance" chi square approximation.
Fourth entry is Cornish-Fisher result.

where $\sigma^2 = \lambda_{11} + \lambda_{22}$, $a_i = \lambda_{ii}/\sigma^2$ and $t = R^2/\sigma^2$. In the three-dimensional situation, we get by the same argument

$$P = P\{a_1x^2 + a_2y^2 + a_3z^2 \leqq t\}, \tag{3}$$

where this time $\sigma^2 = \lambda_{11} + \lambda_{22} + \lambda_{33}$. Similarly, if we leave physical reality, we obtain in k dimensions

$$P = P\left\{\sum_{i=1}^{k} a_i x_i^2 \leqq t\right\} = P\{Q_k \leqq t\} \tag{4}$$

where $\sigma^2 = \sum_1^k \lambda_{ii}$. Now remove the restriction of independence of errors; that is, let the covariance matrix be an arbitrary positive definite matrix. Then there

exists a real non-singular linear transformation [4], $Y \doteq CX$, such that the covariance matrix in the new variables y_i is the unit matrix, and Q_k has the form $\sum_1^k \alpha_i y_i^2$, where the α_i are the roots of the determinantal equation $|A - \alpha\Lambda^{-1}| = 0$, and are all positive, A is the matrix of the coefficients of Q_k considered as a form in the variables x_i, and Λ is the covariance matrix $\{\lambda_{ij}\}$ in these variables. Thus in this paper only (4) is discussed since all other situations can be reduced to it.

3. Exact distribution. Consider the positive definite quadratic form $Q_k = \sum_1^k a_i x_i^2$, where the x_i are normally and independently distributed about zero with unit variance, $\sum a_i = 1$, and $0 < a_i \leqq a_{i+1}$. Denote by $F_k(t)$ the distribution function $F_k(t) = P\{Q_k \leqq t\}$, and by $f_k(t)$ the probability density. Then the Laplace transform $\phi_k(p)$ of $f_k(t)$ is

$$\phi_k(p) = \prod_{j=1}^{k} (1 + 2a_j p)^{-1/2}. \tag{5}$$

From this, $f_k(t)$ and $F_k(t)$ can be obtained in various forms. The authors are including only those which appear most efficient for computing purposes. The following approach was found most useful. Inverting the transform (5) we obtain

$$f_k(t) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{tp}\phi_k(p)\,dp. \tag{6}$$

We now apply Cauchy's theorem to the integrand in (6) taken along the closed contour from $-iR$ to iR along the imaginary axis, from iR to $-R$ along a quarter circle around the origin, from $-R$ to -1 and back along the negative real axis with small clockwise semicircular identations of radius r to avoid the singularities $-\frac{1}{2}a_j$, and from $-R$ back to $-iR$ along a quarter circle around the origin. Letting $R \to \infty$ and $r \to 0$, we obtain

$$f_{2k}(t) = \frac{1}{\pi}\sum_{n=1}^{k} (-1)^{k-n}\int_{-1/2a_{2n-1}}^{-1/2a_{2n}} e^{tp}\phi_{2k}(p)\,dp, \tag{7}$$

$$\begin{aligned} f_{2k+1}(t) = {} & \frac{(-1)^k}{\pi}\int_{-\infty}^{-1/2a_1} e^{tp}\phi_{2k+1}(p)\,dp \\ & + \frac{1}{\pi}\sum_{n=1}^{k} (-1)^{k-n}\int_{-1/2a_{2n}}^{-1/2a_{2n+1}} e^{tp}\phi_{2k+1}(p)\,dp. \end{aligned} \tag{8}$$

We now let $c_j = 1/a_j$, and make the changes of variables

$$p = p_1(x, t) = -\tfrac{1}{2}c_1 - x^2/t, \qquad (-\infty < p < -\tfrac{1}{2}c_1),$$

$$p = p_n(x) = \tfrac{1}{4}(c_{n-1} - c_n)x - \tfrac{1}{4}(c_{n-1} + c_n), \qquad (-\tfrac{1}{2}c_{n-1} < p < \tfrac{1}{2}c_n).$$

For even index we obtain

$$f_{2k}(t) = \frac{(-1)^k}{2\pi}\left\{\prod_{j=1}^{2k}\sqrt{c_j}\right\}\int_{-1}^{1}\sum_{n=1}^{k}(-1)^n G_{2n}(x, t, 2k)\,\frac{dx}{\sqrt{1 - x^2}}, \tag{9}$$

where

$$G_n(x, t, k) = e^{t\,p_n(x)} \prod_{m=1, m\neq n-1, n}^{k} [c_m + 2p_n(x)]^{-1/2}. \tag{10}$$

Integrating (9), we get

$$F_{2k}(t) = 1 + \frac{(-1)^k}{2\pi}\left\{\prod_{j=1}^{2k} \sqrt{c_j}\right\} \int_{-1}^{1} \sum_{n=1}^{k} (-1)^n \frac{G_{2n}(x, t, 2k)}{P_{2n}(x)} \frac{dx}{\sqrt{1-x^2}} \tag{11}$$

Similarly, for odd index,

$$\begin{aligned} f_{2k+1}(t) = {} & \frac{(-1)^k}{2\pi}\left\{\prod_{j=1}^{2k+1} \sqrt{c_j}\right\} \\ & \cdot \int_{-1}^{1} \sum_{n=1}^{k} (-1)^n G_{2n+1}(x, t, 2k+1) \frac{dx}{\sqrt{1-x^2}} + r_{2k+1}(t), \end{aligned} \tag{12}$$

where

$$r_{2k+1}(t) = \frac{(-1)^k}{2\pi}\left\{\prod_{j=1}^{2k+1} \sqrt{c_j}\right\} \left(\frac{t}{2}\right)^{k-\frac{1}{2}} e^{-\frac{1}{2}c_1 t} \int_{-\infty}^{\infty} H(x, t, k) e^{-x^2}\, dx, \tag{13}$$

and

$$H(x, t, k) = \prod_{m=1}^{2k} [x^2 + \tfrac{1}{2}(c_1 - c_{m+1})t]^{-\frac{1}{2}} \tag{14}$$

Integrating (8), we get

$$\begin{aligned} F_{2k+1}(t) = {} & 1 + \frac{(-1)^k}{2\pi}\left\{\prod_{j=1}^{2k+1} \sqrt{c_j}\right\} \\ & \cdot \int_{-1}^{1} \sum_{n=1}^{k} (-1)^n \frac{G_{2n+1}(x, t, 2k+1)}{p_{2n+1}(x)} \frac{dx}{\sqrt{1-x^2}} + R_{2k+1}(t), \end{aligned} \tag{15}$$

where

$$R_{2k+1}(t) = \frac{(-1)^k}{2\pi}\left\{\prod_{j=1}^{2k+1} \sqrt{c_j}\right\} \left(\frac{t}{2}\right)^{k-\frac{1}{2}} e^{-\frac{1}{2}c_1 t} \int_{-\infty}^{\infty} \frac{H(x, t, k)}{p_1(x, t)} e^{-x^2}\, dx. \tag{16}$$

The integrals over the interval $(-1, 1)$ are readily computed using the quadrature formula [16]

$$\int_{-1}^{1} f(x) \frac{dx}{\sqrt{1-x^2}} = \lim_{n\to\infty} \frac{\pi}{n} \sum_{i=1}^{n} f(x_i^{(n)}), \tag{17}$$

where $x_i^{(n)}$ are the zeros of the Tchebycheff polynomials $T_n(x)$ of degree n. Similarly, the zeros $y_i^{(n)}$ and Christoffel numbers $\alpha_i^{(n)}$ of the Hermite polynomials [14] can be used in computing $r_k(t)$ and $R_k(t)$ with the quadrature formula [14], [16]

$$\int_{-\infty}^{\infty} e^{-y^2} f(y)\, dy = \lim_{n\to\infty} \sum_{i=1}^{n} \alpha_i^{(n)} f(y_i^{(n)}). \tag{18}$$

These are usually small unless t is also small, or the two largest coefficients c_1 and c_2 are almost equal. Except under these conditions, they can generally be shown to be negligible by the inequalities

$$|r_{2k+1}(t)|^2 < \frac{1}{2\pi} c_1 \left\{ \prod_{m=1}^{2k} \frac{c_{m+1}}{c_1 - c_{m+1}} \right\} t^{-1} e^{-c_1 t}, \tag{19}$$

$$|R_{2k+1}(t)|^2 < \frac{2}{\pi c_1} \left\{ \prod_{m=1}^{2k} \frac{c_{m+1}}{c_1 - c_{m+1}} \right\} t^{-1} e^{-c_1 t}, \tag{20}$$

which are obtained from (13) and (16) by making use of

$$|H(x, t, k)| \leqq \prod_{m=1}^{2k} \{\tfrac{1}{2}(c_1 - c_{m+1})t\}^{-\frac{1}{2}}, \qquad |p_1(x, t,)| \geqq \tfrac{1}{2} c_1 .$$

For the original two-dimensional problem, we obtain from (9) and (11),

$$f_2(t) = \frac{1}{2\pi} \sqrt{c_1 + c_2}\, e^{-\frac{1}{2}(c_1 + c_2)t} \int_{-1}^{1} e^{\frac{1}{2}(c_1 - c_2)tx} \frac{dx}{\sqrt{1 - x^2}}, \tag{21}$$

$$F_2(t) = 1 - \frac{2}{\pi} \sqrt{c_1 + c_2}\, e^{-\frac{1}{2}(c_1 + c_2)t} \int_{-1}^{1} \frac{e^{\frac{1}{2}(c_1 - c_2)tx}}{(c_1 + c_2) - (c_1 - c_2)x} \frac{dx}{\sqrt{1 - x^2}}, \tag{22}$$

which can be simplified to

$$f_2(t) = \tfrac{1}{2}\sqrt{c_1 + c_2}\, e^{-\frac{1}{2}(c_1 + c_2)t} I_0[\tfrac{1}{2}(c_1 - c_2)t], \tag{23}$$

$$F_2(t) = \frac{2}{\sqrt{c_1 + c_2}} \int_0^{\frac{1}{2}(c_1 + c_2)t} e^{-x} I_0[\sqrt{1/c_2 - 1/c_1}\, x]\, dx, \tag{24}$$

where I_0 is the modified Bessel function of order zero. Although (23) is analytically preferable to (21), (22) is easier to evaluate numerically than (24) except for very small values of t.

The case $k = 3$ applies to the original problem in three dimensions. This time (12) and (15) become

$$\begin{aligned} f_3(t) = {} & \frac{1}{\pi} \sqrt{\frac{c_1 c_2 c_3}{2}}\, e^{-\frac{1}{2}(c_2 + c_1)t} \\ & \cdot \int_{-1}^{1} \frac{e^{-\frac{1}{2}(c_2 - c_1)tx}}{\sqrt{2c_3 - (c_2 + c_1) - (c_2 - c_1)x}} \frac{dx}{\sqrt{1 - x^2}} + r_3(t), \end{aligned} \tag{25}$$

and

$$\begin{aligned} F_3(t) = 1 - {} & \frac{1}{\pi} \sqrt{8 c_1 c_2 c_3}\, e^{-\frac{1}{2}(c_2 + c_1)t} \\ & \cdot \int_{-1}^{1} \frac{e^{-\frac{1}{2}(c_2 - c_1)tx}}{[(c_2 + c_1) + (c_2 - c_1)x]\sqrt{2c_3 - (c_2 + c_1) - (c_2 - c_1)x}} \frac{dx}{\sqrt{1 - x^2}} \\ & + R_3(t) \end{aligned} \tag{26}$$

where

$$r_3(t) = -\frac{1}{\pi}\sqrt{\tfrac{1}{8}c_1c_2c_3}\, t^{\frac{1}{2}} e^{-c_3t/2} \int_{-\infty}^{\infty} \frac{e^{-x^2}\,dx}{\sqrt{[x^2 + \frac{1}{2}(c_3 - c_1)t][x^2 + \frac{1}{2}(c_3 - c_2)t]}}, \tag{27}$$

and

$$\begin{aligned} R_3(t) &= \frac{1}{\pi}\sqrt{\tfrac{1}{8}c_1c_2c_3}\, t^{3/2} e^{-c_3t/2} \\ &\quad\cdot \int_{-\infty}^{\infty} \frac{e^{-x^2}\,dx}{(x^2 + c_3t/2)\sqrt{[x^2 + \frac{1}{2}(c_3 - c_1)t][x^2 + \frac{1}{2}(c_3 - c_2)t]}}. \end{aligned} \tag{28}$$

Numerical evaluation of $f_k(t)$ and $F_k(t)$ becomes more difficult if the constants c_i are almost equal. In that case, however, an as yet unpublished method of Hotelling [8] becomes effective. This will be discussed in the next section. On the other hand, for $f_3(t)$, if two of the constants, say c_j, actually coincide, then the problem simplifies and we obtain as the inverse transform of (5), [5]

$$f_3(t) = \tfrac{1}{2}c_j\sqrt{\frac{c_i}{c_i - c_j}}\, e^{-c_jt/2} \operatorname{erf}\sqrt{\tfrac{1}{2}(c_i - c_j)t}. \tag{29}$$

Hence

$$F_3(t) = I\left(\frac{1}{\sqrt{2}}c_i\,t, -\tfrac{1}{2}\right) - \sqrt{\frac{c_i}{c_i - c_j}}\, e^{-c_jt/2} \operatorname{erf}\sqrt{\tfrac{1}{2}(c_i - c_j)t}, \tag{30}$$

where $I(u, p)$ is the incomplete gamma function as tabulated in [10]. The first entry of each cell in Tables I and II was obtained from the quadrature formulas given above and is correct to four decimal places.

There is an interesting relationship between the distribution of Q_2 and the distribution of the measure of the random set given in [15]. If ${}_a\sigma_{ij} = \sigma_a^2$ for $i = j$ and ${}_a\sigma_{ij} = 0$ for $i \neq j$ and the vector TA mentioned early in the paper is constant, say D, the graph labelled Figure 1 in [15] gives the desired probability if we consider the abscissa values equal to D/σ_a and the ordinate values equal to R/σ_a. Let us now return to our present problem but add the further restriction ${}_p\sigma_{ij} = \sigma_p^2$ for $i = j$ and ${}_p\sigma_{ij} = 0$ for $i \neq j$. Then the probability density of D/σ_p, $h(D/\sigma_p)$, is

$$h\left(\frac{D}{\sigma_p}\right) = \frac{D}{\sigma_p} e^{-\frac{1}{2}(D/\sigma_p)^2}\, d\left(\frac{D}{\sigma_p}\right) \tag{31}$$

and

$$P = P\left\{Q_2 \leqq \frac{R^2}{2(\sigma_a^2 + \sigma_p^2)}\right\} = \int_0^\infty g\left(\frac{R}{\sigma_a}\,\middle|\,\frac{D}{\sigma_a}\right) h\left(\frac{D}{\sigma_p}\right) \tag{32}$$

where $g(R/\sigma_a \mid D/\sigma_a)$ is the probability read from the graph in [15] and the coefficients of Q_2 are now both equal to $\frac{1}{2}$. As an illustration, consider the following four situations: (a) $R/\sigma_a = 2$, $\sigma_p^2/\sigma_a^2 = 3$; (b) $R/\sigma_a = 2$, $\sigma_p^2/\sigma_a^2 = 1$; (c) $R/\sigma_a = 3$, $\sigma_p^2/\sigma_a^2 = 2$; (d) $R/\sigma_a = 3$, $\sigma_p^2/\sigma_a^2 = 1$; then in the table immediately following we get the top entries from Table I, and the bottom entries by numerical integration of (28).

(a)	(b)	(c)	(d)
.3935	.6321	.7769	.8883
.3971	.6328	.7767	.8955

Thus, since only two place accuracy at best could be obtained by reading $g(R/\sigma_a \mid D/\sigma_a)$ from the graph, a rather simple numerical integration yields values extremely close to the exact values.

4. Hotelling's method.[2] Let $2q = Q_k$ and modify Q_k by requiring $\sum a_i = k = 2m$ so that in our cases of special interest $m = 1$ or $\frac{3}{2}$. The a_i are now the ratios of the latent roots of Q_k to k times the trace of the matrix of Q_k where k is rank. Then Hotelling states that the density of q is,

$$f(q) = \frac{q^{m-1}e^{-q}}{\Gamma(m)} \sum_{r=0}^{\infty} b_r L_r(q), \tag{33}$$

where

$$b_r = \frac{r!\,\Gamma(m)}{\Gamma(m+r)} \int_0^\infty f(q) L_r(q)\, dq, \tag{34}$$

and $L_r(q)$ is a Laguerre polynomial defined by

$$L_r(q) = \sum_{t=0}^{r} \binom{r+m-1}{r-t} \frac{(-q)^t}{t!}. \tag{35}$$

Now define

$$u_r = \sum_{j=1}^{k} (a_j - 1)^r. \tag{36}$$

[2] In a letter to one of the authors [8] in November, 1950, Hotelling outlined his method for obtaining the distribution of quadratic forms. This letter was in response to a query regarding a talk Hotelling gave in a seminar attended by one of the authors in Berkeley in 1947. Mention of this research also appears in an abstract by Hotelling in *Ann. Math. Stat.*, Vol. 19 (1948), p. 119.

Then

$$f(q) = \frac{1}{\Gamma(m)} q^{m-1} e^{-q} \left\{ 1 + \frac{u_2}{4}\left[1 - \frac{2q}{m} + \frac{q^2}{m(m+1)}\right] \right.$$

$$- \frac{u_3}{3!}\left[1 - \frac{3q}{m} + \frac{3q^2}{m(m+1)} - \frac{q^3}{m(m+1)(m+2)}\right]$$

$$\tag{37} + \frac{3\left(u_4 + \frac{u_2^2}{4}\right)}{4!}\left[1 - \frac{4q}{m} + \frac{6q^2}{m(m+1)} - \frac{4q^3}{m(m+1)(m+2)} \right.$$

$$\left. + \frac{q^4}{m(m+1)(m+2)(m+3)}\right]$$

$$- \frac{12u_5 + 5u_2 u_3}{5!}\left[1 - \frac{5q}{m} + \frac{10q^2}{m(m+1)} - \frac{10q^3}{m(m+1)(m+2)} \right.$$

$$\left.\left. + \frac{5q^4}{m(m+1)\cdots(m+3)} - \frac{5q^5}{m(m+1)\cdots(m+4)}\right]\right\}$$

+ further terms requiring higher moments of the normal distribution.

Rearranging Hotelling's terms to make optimum use of the Hartley-Pearson Tables [7], we get

$$\tag{38} \begin{aligned} F(t) = \quad & P\{\chi_2^2 \leqq 2t\}\cdot[1 + d_2 - d_3 + d_4 - d_5] \\ & + P\{\chi_4^2 \leqq 2t\}\cdot[-2d_2 + 3d_3 - 4d_4 + 5d_5] \\ & + P\{\chi_6^2 \leqq 2t\}\cdot[d_2 - 3d_3 + 6d_4 - 10d_5] \\ & + P\{\chi_8^2 \leqq 2t\}\cdot[d_3 - 4d_4 + 10d_5] \\ & + P\{\chi_{10}^2 \leqq 2t\}\cdot[d_4 - 5d_5] \\ & + P\{\chi_{12}^2 \leqq 2t\}\cdot[d_5] \end{aligned}$$

where

$$d_2 = \frac{u_2}{4}, \qquad d_3 = \frac{u_3}{6}, \qquad d_4 = \tfrac{1}{8}(u_4 + \tfrac{1}{4}u_2^2), \qquad d_5 = \tfrac{1}{120}(12u_5 + 5u_2 u_3),$$

and χ_n^2 is a chi-square variate with n degrees of freedom. The values obtained by this method using (34) are quite accurate. Using the fixed number of terms in (34), the departure from the exact value depends on the variance of the a_i's. This is noted by a glance at the second entry in each cell of Tables I and II having more than one entry. Thus this method complements the method given in Section 3 precisely in those cases where the most numerical difficulty is experienced; namely, when the variance in the a_i's is small.

5. Approximations. Where a third entry appears in a cell of Tables I and II, it is an approximation obtained in the following way. Let $Q_k = cx_n^2$; this is an approximating device often used in components of variance analysis. Then, equating the first two moments, we get

$$cn = \sum_{i=1}^{k} a_i = 1, \quad c^2 n = \sum_{i=1}^{k} a_i^2.$$

Thus Q_k is approximated by $(\sum_1^k a_i^2)x^2$ where x^2 has $n = 1/\sum_1^k a_i^2$ degrees of freedom. To avoid the interpolation caused by fractional degrees of freedom we can employ the Wilson-Hilferty approximation [17] which states that given a chi-square variate with n degrees of freedom, say χ_n^2, then $(\chi^2/n)^{1/3}$ is approximately normally distributed with mean $(1 - 2/9n)$ and variance $2/9n$; thus we may write

$$P\{Q_k \leqq t\} = P\left\{\left(1 - \frac{2}{9n} + x\sqrt{2/9n}\right)^3 \leqq t\right\} \tag{39}$$

as a modified approximation where x is normally distributed with zero mean and unit variance. Finally we get

$$P\{Q_k \leqq t\} = P\left\{x \leqq \frac{t^{1/3} - (1 - \frac{2}{9}\sum_1^k a_i^2)}{\sqrt{\frac{2}{9}\sum_1^k a_i^2}}\right\}. \tag{40}$$

This result, together with Kelley's Tables [9], was used to obtain the third entry in the cells of the tables wherever they appear.

Where a fourth entry appears in a cell of the tables, it is an approximation obtained from the Cornish-Fisher [3] asymptotic expansion of Q_k in terms of normal variable. This approximation requires the cumulants of Q_k, but these are easy to obtain from the cumulants of the chi-square variate with one degree of freedom by applying the additive properties of cumulants. Computation of the values in Tables I and II is based on all terms in the asymptotic expansion of orders through $1/k^2$.

6. Applications. In discussing applications there is, of course, the obvious one which motivated this paper. As an illustration, assume ${}_a\sigma_{11} = 100$, ${}_a\sigma_{22} = 400$, ${}_p\sigma_{11} = 100$, ${}_p\sigma_{22} = 1400$, and $R = 40$. In this case the usual assumption of circular symmetry is certainly not realistic. Here $a_1 = .1$, $a_2 = .9$, and $t = .8$. Thus the probability of a hit is read as .6159 from Column 5 in Table I. Moreover, Tables I and II make it possible to compare the relative effects of changes in weapon radius with changes in aiming and location errors.

In [2] it is demonstrated that the usual chi-square tests for goodness of fit do not have a limiting chi-square distribution when the maximum likelihood estimates of the parameters are based on the original observations rather than on the cell frequencies. The asymptotic distribution in this situation is that of

$$\sum_{i=1}^{j-s-1} y_i^2 + \sum_{i=j-s}^{j-1} \theta_i y_i^2, \tag{41}$$

where j is the number of cells, s is the number of parameters to be estimated, and the coefficients θ_i are between zero and one and are the roots of a determinantal equation. In the usual "goodness of fit" situation in statistics, distributions rarely contain more than two parameters to be estimated from the data. Thus Tables I and II are singularly appropriate if the number of cells is kept down. In an illustration given in [2],

$$P = P\{x_1^2 + .8x_2^2 + .2x_3^2 \geqq 3.84\} \tag{42}$$

is desired, and $P = .12$ is given as a lower bound. This can be quickly modified so that Table II can be used, for dividing through by two in (38) we get

$$P = P\{.5x_1^2 + .4x_2^2 + .1x_3^2 \geqq 1.92\}. \tag{43}$$

From an Aitken seven point interpolation in the (.5, .4, .1) column in Table II, we get $P = .1344$.

In [1], the limiting distribution of $n\omega^2$ is obtained as the distribution of the quadratic form $Q_\infty = \sum_1^\infty a_i x_i^2$ where $a_i = 1/i^2\pi^2$, and ω^2 is the von Mises criterion for goodness of fit between a sample cumulative distribution function and a specified population distribution function. In [13], it is shown that a simple variant of the ω^2 criterion for the two-sample test has the same limiting distribution. While a table of this distribution is given in [1] it should be possible to use Table II to some advantage, even though this means neglecting all terms from $i = 4$ onwards. Since $\sum_1^\infty a_i = \frac{1}{6} = .1667$ and $\sum_1^3 a_i = 49 / 36\pi^2 = .1379$, a reasonable upper bound should be given by Table II. For example take $t = .046$, $t = .101$, and $t = .405$, then the table in [1] yields .10, .42, and .93 respectively while from Table II we get using

$$P\left\{\frac{x_1^2}{\pi^2} + \frac{x_2^2}{4\pi^2} + \frac{x_3^2}{9\pi^2} \leqq t\right\} = P\left\{\frac{36}{49}x_1^2 + \frac{9}{49}x_2^2 + \frac{4}{49}x_3^2 \leqq \frac{36\pi^2}{49}t\right\} \tag{44}$$

that the probabilities are .28, .54, and .94 respectively. These values are obtained by interpolation and are correct to two places. However, the upper bound is not too sharp when P is small. Also Table II is constructed with t as the argument while the table in [1] has P as the argument and thus may be more useful in some contexts and, of course, less in others.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–211.

[2] H. Chernoff and E. Lehmann, "The use of maximum likelihood estimates in chi square tests for goodness of fit," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 573–578.

[3] E. A. Cornish and R. A. Fisher, "Moments and cumulants in the specification of distributions," *Revue de l'Institut International de Stat.*, Vol. 4 (1937), pp. 307–320.

[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. I, Interscience, New York, 1953.

[5] *Tables of Integral Transforms*, Vol. I, edited by A. Erdélyi, McGraw-Hill, New York, 1954.

[6] J. GURLAND, "Distribution of quadratic forms and ratios of quadratic forms," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 416–427.

[7] H. O. HARTLEY AND E. S. PEARSON, "Tables of the χ^2 integral and of the cumulative Poisson distribution," *Biometrika*, Vol. 37 (1950), pp. 313–325.

[8] H. HOTELLING, Private communication to authors, November, 1950.

[9] T. L. KELLEY, *The Kelley Statistical Tables*, Harvard University Press, 1948.

[10] *Tables of the Incomplete Gamma Function*, edited by Karl Pearson, Biometrika Office, London, 1946.

[11] H. ROBBINS, "The distribution of a definite quadratic form," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 266–270.

[12] H. ROBBINS AND E. J. G. PITMAN, "Applications of the method of mixtures to quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[13] M. ROSENBLATT, "Limit theorems associated with variants of the von Mises statistic," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 617–623.

[14] H. E. SALZER, R. ZUCKER, AND R. CAPUANO, "Table of the zeros and weight factors of the first twenty hermite polynomials," *Journal of Research of the National Bureau of Standards*, Vol. 48, No. 2, February 1952, pp. 111–116.

[15] H. SOLOMON, "Distribution of the measure of a random two-dimensional set," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 650–656.

[16] G. SZEGÖ, *Orthogonal Polynomials*, American Mathematical Society, Colloquium Publications, Vol. 23, New York, 1939.

[17] E. B. WILSON AND M. M. HILFERTY, "The distribution of chi-square" *National Academy of Sciences*, Vol. 17 (1931), pp. 694–698.

URN MODELS OF CORRELATION AND A COMPARISON WITH THE MULTIVARIATE NORMAL INTEGRAL

By J. A. McFadden

U. S. Naval Ordnance Laboratory

Summary. In a special case of Polya's urn scheme, the probability that the first n draws are all of the same color is interpreted as a function of the (single) correlation coefficient. A more general urn model is introduced in which the correlation between pairs of results may differ from pair to pair, and again the probability of consecutive colors is considered. This result is compared with the probability of coincidence in sign under the multivariate normal distribution. The comparison suggests a new approximation for the probability in the multivariate normal case. This approximation appears to be useful only in the Polya case, where the correlations are all equal.

1. Introduction. Consider n correlated random variables $x_1, x_2, \cdots, x_n$. If each variable x_i may assume only the values $+1$ and -1 and either result is equally probable (a priori), then, in terms of the correlation coefficients between pairs (x_i, x_j), what is the probability that all n variables are positive? An example of such a problem is provided by Polya's urn scheme and by a generalization given in Section 3.

A more difficult problem is the following: Consider n correlated *continuous* variables $\xi_1, \xi_2, \cdots, \xi_n$, with each having a mean value of zero and symmetry about the mean. If these variables obey a given distribution law (e.g., the multivariate normal distribution), what is the probability that all n variables are simultaneously positive? This second problem may be reduced, in principle, to the first by associating the signs of the ξ_i with the signs of the x_i; that is,

$$x_i = \begin{cases} 1, & \xi_i \geqq 0, \\ -1, & \xi_i < 0; \end{cases} \qquad i = 1, 2, \cdots, n. \tag{1}$$

The next two sections are concerned with examples of the first problem mentioned above.

2. Polya's urn scheme. Consider the symmetric case of Polya's urn scheme ([1], [2], [3]), in which an urn contains initially a black balls and a red balls. Successive drawings are performed, with replacement, and with the further provision that Δ extra balls are added after each drawing, all of the same color as the ball most recently drawn. Δ may be negative, but it must obey the inequality,

$$a + (n-1)\Delta \geqq 0, \tag{2}$$

where n is the total number of draws, in order that neither color may be overdrawn.

Received November 1, 1954.

The probability of drawing a black ball in the first trial is, of course, $a/2a = \frac{1}{2}$. The probability of drawing two black balls in the first two trials is $a(a + \Delta) / 2a(2a + \Delta)$. The probability of drawing n black balls in the first n trials is

(3) $$P_n = (a/\Delta)_n \,/\, (2a/\Delta)_n ,$$

where $(\alpha)_n = \alpha(\alpha + 1)(\alpha + 2) \cdots (\alpha + n - 1)$, and $(\alpha)_0 = 1$.

Let $x_i = +1$ if the ith draw is black, and let $x_i = -1$ if the ith draw is red. Then Polya has shown ([3], p. 140) that the correlation coefficient between x_i and x_j $(i \neq j)$ is

(4) $$r = \Delta \,/\, (2a + \Delta).$$

The result is the same for all possible pairs (i, j). As Δ varies from $-a \,/\, (n - 1)$ to ∞, r varies from $-1 \,/\, (2n - 3)$ to 1.

Equation (4) may be verified easily for the case $i = 1, j = 2$. Since the mean values $E(x_i)$ are all zero and the variances $E(x_i^2)$ are all unity, the correlation coefficient between x_i and x_j is simply the expectation $E(x_i x_j)$. For the first two draws, there are four possibilities: $(+1, +1)$, $(+1, -1)$, $(-1, +1)$, and $(-1, -1)$. Then

$$r = E(x_1 x_2) = 2\,\frac{a(a + \Delta)}{2a(2a + \Delta)} - 2\,\frac{a^2}{2a(2a + \Delta)} = \frac{\Delta}{2a + \Delta},$$

which agrees with (4). The same procedure may be carried out for other pairs (i, j).

a/Δ may now be eliminated from (3) by equation (4); then, in terms of the correlation alone, the probability that all the x_i are equal is given by

(5) $$P_n = ([1 - r] \,/\, 2r)_n \,/\, ([1 - r] \,/\, r)_n .$$

At this point the integers a and Δ may be forgotten. Let r assume *any* value in the range from $-1 \,/\, (2n - 3)$ to 1—not only the fractional values given by equation (4).

P_n may be expressed in terms of the beta function, as follows:

(6) $$P_n = \frac{B(n + [1 - r] \,/\, 2r, [1 - r] \,/\, 2r)}{B([1 - r] \,/\, 2r, [1 - r] \,/\, 2r)},$$

or, equivalently, as a terminating hypergeometric series:

$$P_n = 2^{-n} F(-n/2, [1 - n] \,/\, 2; 1/2r; 1)$$

(7) $$= 2^{-n}\left\{1 + \frac{n(n - 1)}{2(1)}\,r + \frac{n(n - 1)(n - 2)(n - 3)}{2 \cdot 4(1)(1 + 2r)}\,r^2 + \cdots\right\}.$$

[The identity between (6) and (7) follows from the theorem on $F(a, b; c; 1)$—see [4], p. 282—and from the multiplication theorem for gamma functions, [4], p. 240. See also [4], p. 262, problem 37.]

Note that if $r \to 0$, the probability (7) approaches 2^{-n}, which is the usual result for a sequence of Bernoulli trials when the individual probabilities are $\frac{1}{2}$.

If $r \to 1$, the expression in braces in (7) becomes the binomial series for $\frac{1}{2}[(1+1)^n + (1-1)^n]$ and therefore $P_n \to \frac{1}{2}$. This is the case of perfect coherence, which leaves only two possibilities: all black or all red.

3. A generalized urn scheme. It was noted in Section 2 that the Polya scheme yields a correlation matrix in which all elements except those on the main diagonal have the value (4). The following model exhibits a general correlation matrix.

As before, the urn contains initially a black balls and a red balls. In contrast with the single addition parameter Δ, this scheme makes use of a *matrix* of elements Δ_{ij}. One ball is drawn and replaced, and Δ_{12} balls are added of the color drawn. Again one ball is drawn and replaced; then Δ_{13} are added of the first color drawn and Δ_{23} of the second color drawn. After the $(k-1)$th draw (and replacement), Δ_{1k} are added of the first color drawn, Δ_{2k} of the second, etc., and $\Delta_{k-1,k}$ of the $(k-1)$th color drawn.

To simplify the algebra, let

(8) $$D_{mk} = \sum_{i=m+1}^{k} \Delta_{mi}, \qquad m = 1, 2, \cdots, k-1;$$

thus, immediately preceding the kth draw, D_{mk} is the total number of balls which have been added up to that time *because of* the mth draw. Some of the Δ's may be negative, but to prevent overdrawing they must obey the inequality,

(9) $$a + \sum_{m=1}^{k-1} D_{mk} \geqq 0,$$

for all integral k between 2 and n, inclusive. (n is again the total number of draws.)

The probabilities of the sequences black-black and black-red in the first two draws are, respectively,

(10) $$P_{++} = \frac{a(a+D_{12})}{2a(2a+D_{12})}, \qquad P_{+-} = \frac{a(a)}{2a(2a+D_{12})}.$$

By symmetry, $P_{--} = P_{++}$ and $P_{-+} = P_{+-}$, as in the Polya scheme. For three draws the probabilities are

(11)
$$P_{+++} = \frac{a(a+D_{12})(a+D_{13}+D_{23})}{2a(2a+D_{12})(2a+D_{13}+D_{23})},$$

$$P_{++-} = \frac{a(a+D_{12})(a)}{2a(2a+D_{12})(2a+D_{13}+D_{23})},$$

$$P_{+-+} = \frac{a(a)(a+D_{13})}{2a(2a+D_{12})(2a+D_{12}+D_{23})},$$

$$P_{+--} = \frac{a(a)(a+D_{23})}{2a(2a+D_{12})(2a+D_{13}+D_{23})},$$

and the other four may be obtained by symmetry.

What are the correlation coefficients in this scheme? Again let $x_i = 1$ or -1 if the ith draw is black or red, respectively, so that $r_{ij} = E(x_i x_j)$. For the first two draws,

$$r_{12} = E(x_1 x_2) = 2(P_{++} - P_{+-}) = D_{12} / (2a + D_{12}). \tag{12}$$

The last equality follows from the substitution of (10). Equation (12) conforms with (4), since for two draws the two urn schemes are identical.

For the first three draws, by (10) and (11),

$$\begin{aligned} r_{13} &= 2(P_{+++} - P_{++-} - P_{+--} + P_{+-+}) \\ &= 2(P_{+++} - P_{++-}) + 2(P_{+-+} - P_{+--}) \\ &= 2P_{++} \frac{D_{13} + D_{23}}{2a + D_{13} + D_{23}} + 2P_{+-} \frac{D_{13} - D_{23}}{2a + D_{13} + D_{23}} \\ &= 2 \frac{D_{13}(P_{++} + P_{+-}) + D_{23}(P_{++} - P_{+-})}{2a + D_{13} + D_{23}} \\ &= \frac{D_{13} + r_{12} D_{23}}{2a + D_{13} + D_{23}}, \end{aligned} \tag{13}$$

and by a similar calculation,

$$r_{23} = \frac{r_{12} D_{13} + D_{23}}{2a + D_{13} + D_{23}}. \tag{14}$$

The above method is easily generalized for the first n draws. The result is

$$r_{in} = \frac{\sum_{j=1}^{n-1} r_{ij} D_{jn}}{2a + \sum_{j=1}^{n-1} D_{jn}}, \qquad i = 1, 2, \cdots, n-1, \tag{15}$$

where $r_{ii} = 1$ for all i. Notice that if all the correlation coefficients are known for the first $(n-1)$ draws, then equation (15) gives the remaining coefficients necessary to correlate the nth draw.

The next quantity to be calculated is the ratio $2P_n / P_{n-1}$, where P_n is again the probability of drawing n black balls in the first n trials. ($P_1 = P_+ = \frac{1}{2}$, $P_2 = P_{++}$, $P_3 = P_{+++}$, etc.) By the first of equations (10),

$$2 \frac{P_{++}}{P_+} = \frac{2(a + D_{12})}{2a + D_{12}} = 1 + \frac{D_{12}}{2a + D_{12}}. \tag{16}$$

By equations (10) and (11),

$$2 \frac{P_{+++}}{P_{++}} = 1 + \frac{D_{13} + D_{23}}{2a + D_{13} + D_{23}}. \tag{17}$$

Similarly, for n draws,

$$(18)\qquad 2\frac{P_n}{P_{n-1}} = 1 + \frac{\sum_{j=1}^{n-1} D_{jn}}{2a + \sum_{j=1}^{n-1} D_{jn}}, \qquad n = 2, 3, \cdots.$$

The next objective is to express the ratios (18) as functions *of the correlation coefficients alone.* A new variable is introduced:

$$(19)\qquad G_{kn} = \frac{D_{kn}}{2a + \sum_{j=1}^{n-1} D_{jn}}, \qquad k = 1, 2, \cdots, n-1.$$

Then equations (15) may be written

$$(20)\qquad r_{in} = \sum_{j=1}^{n-1} r_{ij} G_{jn}, \qquad i = 1, 2, \cdots, n-1,$$

and equation (18) may be written in the form

$$(21)\qquad -1 = \sum_{j=1}^{n-1} G_{jn} - 2P_n / P_{n-1}.$$

Now equations (20) and (21) constitute n equations in the n unknowns, $G_{1n}, G_{2n}, \cdots, G_{n-1,n}$, and $2P_n / P_{n-1}$. The equations may be solved directly for the last quantity; then a simple manipulation of the determinants yields the result,

$$(22)\qquad 2\frac{P_n}{P_{n-1}} = \frac{\begin{vmatrix} 1 + r_{1n} & r_{12} + r_{1n} & \cdots & r_{1,n-1} + r_{1n} \\ r_{12} + r_{2n} & 1 + r_{2n} & \cdots & r_{2,n-1} + r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ r_{1,n-1} + r_{n-1,n} & r_{2,n-1} + r_{n-1,n} & \cdots & 1 + r_{n-1,n} \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} & \cdots & r_{1,n-1} \\ r_{12} & 1 & \cdots & r_{2,n-1} \\ \cdot & \cdot & \cdot & \cdot \\ r_{1,n-1} & r_{2,n-1} & \cdots & 1 \end{vmatrix}}, \quad n = 2, 3, \cdots.$$

The denominator is simply the determinant of the $(n-1)$-variate correlation matrix.

As in the case of the Polya scheme, the correlation coefficients may now be regarded as continuous rather than discrete. These coefficients may assume any values between -1 and 1 which do not violate (9), i.e., which do not lead to negative probabilities. By comparison with equation (18), it follows that the inequality (9) may be rewritten

$$(23)\qquad P_k / P_{k-1} \geqq 0,$$

for $k = 2, 3, \cdots, n$, where P_k / P_{k-1} is given by (22).

Finally, by induction, the probability P_n is given by

$$P_n = 2^{-n} \frac{(1 + r_{12}) \begin{vmatrix} 1 + r_{13} & r_{12} + r_{13} \\ r_{12} + r_{23} & 1 + r_{23} \end{vmatrix}}{1 \begin{vmatrix} 1 & r_{12} \\ r_{12} & 1 \end{vmatrix}}$$

(24)
$$\cdots \frac{\begin{vmatrix} 1 + r_{1n} & r_{12} + r_{1n} & \cdots & r_{1,n-1} + r_{1n} \\ r_{12} + r_{2n} & 1 + r_{2n} & \cdots & r_{2,n-1} + r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ r_{1,n-1} + r_{n-1,n} & r_{2,n-1} + r_{n-1,n} & \cdots & 1 + r_{n-1,n} \end{vmatrix}}{\begin{vmatrix} 1 & r_{12} & \cdots & r_{1,n-1} \\ r_{12} & 1 & \cdots & r_{2,n-1} \\ \cdot & \cdot & \cdot & \cdot \\ r_{1,n-1} & r_{2,n-1} & \cdots & 1 \end{vmatrix}}, \quad n = 2, 3, \cdots.$$

When all the coefficients r_{ij} are equal for $i, j = 1, 2, \cdots, n$ $(i \neq j)$, equation (24) reduces to the Polya result (5).

When $n = 2$ and 3 the result (24) becomes simply

$$P_2 = \tfrac{1}{4}(1 + r_{12}) \tag{25}$$

$$P_3 = \tfrac{1}{8}(1 + r_{12} + r_{13} + r_{23}). \tag{26}$$

For higher values of n the complete expansion of (24) is quite complicated. However P_n may be expanded in a power series in the r's. To second order, for $n = 4$,

$$P_4 = \tfrac{1}{16}(1 + r_{12} + r_{13} + r_{23} + r_{14} + r_{24} + r_{34} + r_{12}r_{34} + r_{13}r_{24} + r_{14}r_{23}) + O(r^3). \tag{27}$$

By induction, it can be shown that, for general n,

$$P_n = 2^{-n}\left[1 + \sum_{j>i\geq 1}^{n} r_{ij} + \sum_{l>k>j>i\geq 1}^{n} (r_{ij}r_{kl} + r_{ik}r_{jl} + r_{il}r_{jk}) + O(r^3)\right]. \tag{28}$$

When all the coefficients r_{ij} are equal $(i \neq j)$, the number of first-order terms in (28) is the binomial coefficient $\binom{n}{2}$. The number of second-order terms is $3\binom{n}{4}$; therefore this special case of (28) checks (to second order) with the result (7) of the Polya model.

If P_n is expanded to a higher order, then the series is no longer symmetric with respect to interchange of the variables $x_1, x_2, \cdots, x_n$.

4. The multivariate normal distribution. The following example belongs to the second type of problem given in the introduction.

Suppose that $\xi_1, \xi_2, \cdots, \xi_n$ obey the multivariate normal distribution law with correlation matrix

$$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{pmatrix}. \tag{29}$$

If all the mean values $E(\xi_i)$ are zero, what is the probability P_n that all n variables are simultaneously positive?

As stated in the introduction, this question may be reduced to the corresponding problem in the discrete variables $x_1, x_2, \cdots, x_n$ by means of equation (1); however, it must be remembered that the correlation r_{ij} between x_i and x_j is not the same as the correlation ρ_{ij} between ξ_i and ξ_j.

Various writers have investigated the probability P_n for the multivariate normal distribution. For $n = 2$ there is the Stieltjes-Sheppard result [5],

$$P_2 = \tfrac{1}{4}\left(1 + \frac{2}{\pi}\sin^{-1}\rho_{12}\right). \tag{30}$$

For $n = 3$ the result is (see Kendall [7] and David [8]):

$$P_3 = \tfrac{1}{8}\left[1 + \frac{2}{\pi}(\sin^{-1}\rho_{12} + \sin^{-1}\rho_{13} + \sin^{-1}\rho_{23})\right]. \tag{31}$$

For $n > 3$ no solution has been given in closed form, but there exists the infinite series of Aitken, Kendall ([6], [7]), and Moran [9]. For $n = 4$, their series may be written, to second order in the ρ_{ij},

$$P_4 = \tfrac{1}{16}\left[1 + \frac{2}{\pi}\sum_{j>i\geq 1}^{4}\sin^{-1}\rho_{ij} + \frac{4}{\pi^2}\{\rho_{12}\rho_{34} + \rho_{13}\rho_{24} + \rho_{14}\rho_{23} + O(\rho^3)\}\right]. \tag{32}$$

For general n, the probability is

$$\begin{aligned} P_n = 2^{-n}\Bigg[1 + \frac{2}{\pi}\sum_{j>i\geq 1}^{n}\sin^{-1}\rho_{ij} \\ + \frac{4}{\pi^2}\sum_{l>k>j>i\geq 1}^{n}(\rho_{ij}\rho_{kl} + \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk}) + O(\rho^3)\Bigg]. \end{aligned} \tag{33}$$

(This result will be derived in Section 6.)

Equation (33) may be compared with the corresponding probability (28) for the generalized urn scheme. Note that under the transformation,

$$r_{ij} = \frac{2}{\pi}\sin^{-1}\rho_{ij}, \tag{34}$$

the two expressions (28) and (33) agree to second order. [Note that r_{ij}, given by (34), is actually the correlation between the discrete variables x_i and x_j, given by equation (1), when the ξ's are normal. See equation (41).] This agree-

ment suggests that the substitution of (34) into the *closed form* (24) of the result for the generalized urn scheme might provide an approximation for P_n in the multivariate normal case, to be used in place of the poorly converging series of Aitken, Kendall, and Moran. (See David's remarks [8] on convergence.)

For $n > 3$ and arbitrary ρ_{ij}, the agreement between the two power series does not extend beyond the second-order terms.

5. Numerical results. The approximation indicated above has been tested by a comparison with several known results for the multivariate normal integral.

When all the ρ_{ij} have the same value, defined by ρ, then r is obtained from equation (34); then this value of r is used in the closed expression (5) of the Polya scheme.

When $n = 2$ or 3, the result obtained from (5) is exact for all values of ρ; that is, equation (30) and the special case of (31) follow immediately.

When $\rho = 0$, (5) gives $P_n = 2^{-n}$; when $\rho = 1$, $P_n = \frac{1}{2}$. (See the explanation at the end of Section 2.) It appears, therefore, that when $\rho = 0$ or 1 the results are exact for all values of n.

When $\rho = \frac{1}{2}$, then $r = \frac{1}{2}$ and equation (5) gives $P_n = 1 / (n + 1)$. This result is also exact for all n. {See Ruben [10], p. 214, equation (70). In fact, Ruben's (70) holds for a more general class of distributions, as shown by Foster and Stuart [11], p. 22.}

When $1/\rho = 2, 3, \cdots, 12$, the results obtained from equation (5) may be compared with those of Ruben ([10], pp. 222–223). For the case $n = 4$, the comparison is shown in Table I. (Ruben's values have been rounded off to seven decimal places.)

The best agreement in Table I occurs for small ρ, as one might have predicted after a comparison of the corresponding power series.

For a given value of ρ, the approximation grows steadily worse as n increases. A comparison for $\rho = \frac{1}{4}$ is shown in Table II.

TABLE I
($n = 4$)

$1/\rho$	P_4 [from (5)]	Ruben's $\Phi_4(1/\rho)$
2	0.20000 00	0.20000 00
3	0.14975 57	0.14973 77
4	0.12649 38	0.12647 92
5	0.11302 30	0.11301 25
6	0.10423 15	0.10422 40
7	0.09804 22	0.09803 67
8	0.09344 92	0.09344 51
9	0.08990 58	0.08990 27
10	0.08708 94	0.08708 71
11	0.08479 73	0.08479 5
12	0.08289 56	0.08289 4

TABLE II

($\rho = \frac{1}{4}$)

n	P_n [from (5)]	Ruben's $\Phi_n(4)$
2	0.29021 53	0.29021 53
3	0.18532 30	0.18532 30
4	0.12649 38	0.12647 92
5	0.09069 62	0.09065 98
6	0.06754 16	0.06748 27
7	0.05183 56	0.05175 69
8	0.04076 86	0.04067 37
9	0.03272 29	0.03261 57
10	0.02671 93	0.02660 32

TABLE III

ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	P_4 [from (24)]	Plackett's Φ_4^0
$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0.13393 0.13194 0.13194 0.13393	0.13333
$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0.16369 0.16369 0.16369 0.16369	0.16667
$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0.15000 0.15278 0.14881 0.14881	0.15000

It appears from Tables I and II that the Polya urn approximation might be useful in many problems, at least where ρ is not greater than $\frac{1}{2}$ and where n is not much greater than 4. Formulas (5) and (34) are certainly more easily applicable than Ruben's integral recursion formulas or an interpolation in Ruben's table.

In the general case of unequal ρ_{ij}, the results are much less satisfactory. This fact can be illustrated by a comparison with several exact values given by Plackett ([12], p. 360) for the quadrivariate case. The comparison is shown in Table III.

To obtain P_4, one substitutes the values of r_{ij} from (34) into the closed expression (24). When $n > 3$, (24) is not symmetric with respect to interchange of the indices; therefore different results are possible. The four values of P_4 given in Table III (for each correlation matrix) are those obtained when x_4, x_3, x_2, and x_1, respectively, are considered as the fourth draw from the urn. Without doubt the lack of symmetry in (24) is partly responsible for the poor agreement.

It appears from Table III (and from other comparisons with Plackett's figures) that the generalized urn approximation is not a satisfactory method for computing the general multivariate normal integral. The only possible exceptions would be those situations in which the ρ_{ij} are very small or are all nearly equal to each other, i.e., approaching the Polya case.

Recently Plackett [12] has given a numerical method for evaluating the quadrivariate normal integral. It involves more labor than the urn scheme but yields considerably greater accuracy.

6. General remarks. This paper will be concluded by a discussion of the general problem described in the introduction, in which no specific model or distribution law is assumed.

Consider again the n mutually interacting (discrete) random variables $x_1, x_2, \cdots, x_n$. Let the *observed values* of these variables (in a given experiment) be $y_1, y_2, \cdots, y_n$, where $y_i = \pm 1$. Then there are 2^n possible combinations for the n results y_i, and each result has the probability

$$P(x_1 = y_1, x_2 = y_2, \cdots, x_n = y_n). \tag{35}$$

For any given distribution law there are 2^n product moments, i.e., the expected values $E(1)$, $E(x_j)$, $E(x_jx_k)$, $E(x_jx_kx_l)$, $\cdots$, $E(x_1x_2 \cdots x_n)$. All other moments degenerate to one of these, since $x_i^2 = 1$[e.g., $E(x_1^2) = E(1)$; $E(x_1^2x_2) = E(x_2)$].

The 2^n moments may be expressed in terms of the 2^n probabilities, as follows:

$$\begin{aligned}
E(1) &= \sum_{y_i} 1 \cdot P(x_1 = y_1, x_2 = y_2, \cdots, x_n = y_n), \\
E(x_j) &= \sum_{y_i} y_j P(x_1 = y_1, x_2 = y_2, \cdots, x_n = y_n), \quad j = 1, 2, \cdots, n, \\
E(x_jx_k) &= \sum_{y_i} y_jy_k P(x_1 = y_1, x_2 = y_2, \cdots, x_n = y_n), \\
&\qquad\qquad j, k = 1, 2, \cdots, n(j < k), \\
&\cdots\cdots\cdots\cdots \\
E(x_1x_2 \cdots x_n) &= \sum_{y_i} y_1y_2 \cdots y_n P(x_1 = y_1, x_2 = y_2, \cdots, x_n = y_n),
\end{aligned} \tag{36}$$

where the sums are taken over all combinations $y_1 = \pm 1$, $y_2 = \pm 1, \cdots, y_n = \pm 1$.

If all the equations (36) are added together, then all the probabilities cancel except $P(x_1 = x_2 = \cdots = x_n = 1)$, which was previously called P_n. Then

$$P_n = 2^{-n}\left[1 + \sum_{i=1}^{n} E(x_i) + \sum_{j>i\geq 1}^{n} E(x_ix_j) + \cdots + E(x_1x_2 \cdots x_n)\right]. \tag{37}$$

Now suppose that the symmetry of the distribution law is such that all product moments *of odd order* are zero, i.e.,

$$\begin{aligned}
E(x_1) &= E(x_2) = \cdots = E(x_n) = 0, \\
E(x_1x_2x_3) &= E(x_1x_2x_4) = \cdots = E(x_{n-2}x_{n-1}x_n) = 0,
\end{aligned} \tag{38}$$

etc. Then (37) becomes

$$(39)\quad P_n = 2^{-n}\left[1 + \sum_{j>i\geq 1}^{n} E(x_i x_j) + \sum_{l>k>j>i\geq 1}^{n} E(x_i x_j x_k x_l) + \cdots\right],$$

ending with a product moment of order n if n is even, or with moments of order $(n-1)$ if n is odd.

It is now evident that P_2 and P_3 in equations (25) and (26) could be obtained directly from the general formula (39) by the substitution of $n = 2$ or 3 and of the definition $E(x_i x_j) = r_{ij}$. In other words, it is only for $n > 3$ that a specific urn model must be assumed, and this specialization is reflected in the values of the higher-order moments $E(x_i x_j x_k x_l)$, etc. in (39).

On the other hand, suppose ξ_i are *continuous* random variables obeying a given distribution law with symmetry as in (38). Then for the calculation of P_2 and P_3, $E(x_i x_j)$ must be obtained as a function of the parameters of the original distribution. For P_4 and P_5, $E(x_i x_j x_k x_l)$ must be obtained, etc., and all other P's will follow two at a time from the higher moments.

It is now possible to derive equations (31) and (33). Assume that P_2 is given by (30). By matching (30) with (39) when $n = 2$, it follows that

$$(40)\quad E(x_1 x_2) = \frac{2}{\pi} \sin^{-1} \rho_{12}.$$

Then by the symmetry of the normal distribution,

$$(41)\quad E(x_i x_j) = \frac{2}{\pi} \sin^{-1} \rho_{ij},$$

for all i and j, $i \neq j$, and equation (31) follows by the substitution of the moments (41) into the general expression (39) with $n = 3$.

Now assume that P_4 is given correctly by equation (32). Then (32) may be matched with (39) when $n = 4$, with the aid of (41), and the result is given by

$$(42)\quad E(x_1 x_2 x_3 x_4) = \frac{4}{\pi^2} (\rho_{12}\rho_{34} + \rho_{13}\rho_{24} + \rho_{14}\rho_{23}) + O(\rho^3).$$

Then, by symmetry, the general fourth-order product moment is

$$(43)\quad E(x_i x_j x_k x_l) = \frac{4}{\pi^2} (\rho_{ij}\rho_{kl} + \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk}) + O(\rho^3), \qquad i < j < k < l,$$

and, since all higher-order moments are of higher order in the ρ_{ij}, equation (33) follows. The last operation is the substitution of the moments (41) and (43) into the general expression (39).

This process is equivalent to a method used by David [8] to obtain P_{n+1} from P_n when n is even.

Acknowledgment. The author is grateful to Dr. H. E. Ellingson, Mr. L. D. Krider, and Dr. A. Van Tuyl for many helpful discussions on this problem.

REFERENCES

[1] W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1950, pp. 82–83.

[2] F. Eggenberger and G. Polya, "Über die Statistik verketteter Vorgänge," *Z. angew. Math. Mech.*, Vol. 3 (1923), pp. 279–289.

[3] G. Polya, "Sur quelques points de la théorie des probabilités," *Ann. Inst. H. Poincaré*, Vol. 1 (1930), pp. 117–161.

[4] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, 1946, American Edition.

[5] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 290.

[6] M. G. Kendall, "Proof of relations connected with the tetrachoric series and its generalization," *Biometrika*, Vol. 32 (1941), pp. 196–198.

[7] M. G. Kendall, *Contributions to the Study of Oscillatory Time-Series*, Cambridge University Press, 1946, pp. 59–62.

[8] F. N. David, "A note on the evaluation of the multivariate normal integral," *Biometrika*, Vol. 40 (1953), pp. 458–459.

[9] P. A. P. Moran, "Rank correlation and product-moment correlation," *Biometrika*, Vol. 35 (1948), pp. 203–206.

[10] H. Ruben, "On the moments of order statistics in samples from normal populations," *Biometrika*, Vol. 41 (1954), pp. 200–227.

[11] F. G. Foster and A. Stuart, "Distribution-free tests in time-series based on the breaking of records," *J. Roy. Statist. Soc.*, Ser. B, Vol. 16 (1954), pp. 1–22.

[12] R. L. Plackett, "A reduction formula for normal multivariate integrals," *Biometrika*, Vol. 41 (1954), pp. 351–360.

STATISTICS AND SUBFIELDS[1]

By R. R. Bahadur

The University of Chicago

1. Introduction and summary. Let $(X, \mathbf{S}, \mu)$ be a probability measure space: X is set of points x, $\mathbf{S}$ is a field of subsets of X, and μ is a countably additive measure on $\mathbf{S}$ with $\mu(X) = 1$.[2] A *subfield* is a field $\mathbf{S}_0$ of subsets of X such that $\mathbf{S}_0 \subseteq \mathbf{S}$, that is, each $\mathbf{S}_0$-measurable set is also $\mathbf{S}$-measurable. A *statistic* is a function defined on X. There is no a priori restriction on the class of statistics; in particular, statistics are not necessarily real-valued, and a real-valued statistic is not necessarily an $\mathbf{S}$-measurable function. For any statistic f, let $\mathbf{S}_f$ denote the class of all sets which are $\mathbf{S}$-measurable and of the form $f^{-1}(B)$, where B is a subset of the range of f. The class $\mathbf{S}_f$ is clearly a subfield, and is called the subfield *induced by f*.

The induced subfield $\mathbf{S}_f$ plays a central role in the study of a statistic f, for the following reason. The probabilist or mathematical statistician is usually concerned not with the statistic f as such, but rather with the class of random variables (i.e., real-valued $\mathbf{S}$-measurable functions) which depend on x only through f, and, as is easily seen, this class of random variables is exactly the class of real-valued $\mathbf{S}_f$-measurable functions. In case the given statistic f is a random variable (and therefore itself an object of study), the argument just given continues to apply, because in this case f is necessarily an $\mathbf{S}_f$-measurable function.

This paper discusses certain measure-theoretic problems concerning the relations between subfields, subfields of the apparently special form $\mathbf{S}_f$, and statistics. The main problems, as also the main conclusions, are described in the following paragraphs. Most of the conclusions of the paper are valid only in the case when $(X, \mathbf{S})$ is (or may be taken to be) a euclidean sample space, that is, X is a Borel set of the m-dimensional euclidean space $(1 \leqq m \leqq \infty)$, and $\mathbf{S}$ is the field of Borel sets of X. It is assumed henceforth that this is the case under consideration.

There are two main problems. The first is whether every subfield is inducible by a statistic. This problem is discussed (in a more general setting) in [2], and the conclusions of the present paper complement those of [2].

It is shown here that every subfield is inducible by a statistic if and only if the sample space is discrete, that is to say, X is a countable set and $\mathbf{S}$ is the class of all subsets of X (Theorem 1). This result is, however, not quite relevant to situations where the natural equivalence relation between subfields is not identity but approximability to within sets of μ-measure zero. The equivalence relation

Received May 24, 1954.

[1] This work was supported in part by the Office of Naval Research under Contract N6onr-271, T.O. X1, Project 042-034.

[2] This paper uses some of the notation and terminology of the first part of [1]. In particular, all fields considered are understood to be countably additive.

referred to is defined as follows. A subfield $\mathbf{S}_1$ is a *contraction* of a subfield $\mathbf{S}_2$ if corresponding to each real-valued $\mathbf{S}_1$-measurable function f_1 there exists an $\mathbf{S}_2$-measurable function f_2 such that $f_2(x) = f_1(x)$ except on a set of μ-measure zero; we then write $\mathbf{S}_1 \subseteq \mathbf{S}_2$ $[\mathbf{S}, \mu]$. The subfields $\mathbf{S}_1$ and $\mathbf{S}_2$ are *equivalent* if each is a contraction of the other; we then write $\mathbf{S}_1 = \mathbf{S}_2$ $[\mathbf{S}, \mu]$. It is shown that, in fact, corresponding to any subfield $\mathbf{S}_0$ there exists an f such that $\mathbf{S}_f$ is equivalent to $\mathbf{S}_0$, and that this f may be taken to be a random variable (Theorem 2).

In the literature the notion of contraction (and the derived notion of equivalence) has been defined for statistics in two ways, which are here called contraction and functional contraction. A statistic f is a *contraction* of a statistic g if $\mathbf{S}_f$ is a contraction of $\mathbf{S}_g$ (that is, $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$); f is a *functional contraction* of g (written $f \subseteq g$ $[\mathbf{S}, \mu]$) if there exists a function h on the range of g into that of f, and an $\mathbf{S}$-measurable set N with $\mu(N) = 0$, such that $f(x) = h(g(x))$ for x in $X - N$. (Cf. [3], [4].) It seems to the writer that for most (possibly all) technical purposes the relevant concept is contraction as just defined (cf. Lemmas 7.1 and 3.2 of [1]). However, functional contraction has simpler interpretations and greater intuitive appeal.

The second problem is the exact relation between contraction and functional contraction. It is shown that, in general, functional contraction does not imply contraction (Example 1), and also that contraction does not imply functional contraction (Example 2). If, however, both f and g are random variables, then $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$ if and only if $f \subseteq g$ $[\mathbf{S}, \mu]$ (Theorem 3). It follows, in particular, that if the sample space is discrete, then contraction coincides with functional contraction.

The problems described above arose in connection with the theory of sufficiency, and the results have applications in that theory. It follows, for example (assuming that the sample space is euclidean and that the set of alternative distributions of the sample point is a dominated set), that if f is a necessary and sufficient statistic, then $\mathbf{S}_f$ is a necessary and sufficient subfield (Corollary 2).

The following are some general conclusions bearing on mathematical models for studies such as [1]. (a) The notion of a subfield, while certainly no less general than that of a statistic, is in fact no more general. (b) There is no loss of generality, or other disadvantage, in defining a statistic to be a random variable. On the contrary, admission of nonmeasurable functions to the discussion leads to inconsistencies between extension and functional extension—this seems undesirable. (c) If f is a random variable, it is immaterial whether f is regarded as a statistic or as a Borel-measurable transformation (cf. [1], p. 431). These satisfactory conclusions do not necessarily hold for an arbitrary space $(X, \mathbf{S}, \mu)$. An example given in [2] shows that at least (a) and (c) are not valid in the general case.

2. Theorems. Let R be the real line, and $\mathbf{R}$ be the class of Borel sets of R. In general, we shall denote the n-dimensional euclidean space by R^n and the class of Borel sets of R^n by $\mathbf{R}^n$ $(1 \leqq n \leqq \infty)$. The following well-known result (cf. [5], pp. 159–160) is stated here as a lemma for convenience of reference.

LEMMA 1. *There exists a one to one function α_n on R^n onto R such that $A \in \mathbf{R}^n$ implies $\alpha_n(A) \in \mathbf{R}$ and $B \in \mathbf{R}$ implies $\alpha_n^{-1}(B) \in \mathbf{R}^n$ $(1 \leqq n \leqq \infty)$.*

If f is a statistic on X into a space Y, and $\mathbf{C}$ is a class of subsets of Y, then $f^{-1}(\mathbf{C})$ denotes the class of all sets of X which are of the form $f^{-1}(B)$ with $B \in \mathbf{C}$. Clearly, $f^{-1}(\mathbf{C})$ is a field if and only if $\mathbf{C}$ is a field. A function f on X into R^n is said to be $\mathbf{S}$-measurable if $f^{-1}(\mathbf{R}^n)$ is a subfield of $\mathbf{S}$.

In this section and the following one, a number of results involving $\mathbf{S}$-measurable functions (specifically, Lemmas 2, 4, 5, 6, 8, and 9; Theorems 2 and 3; Corollaries 1 and 3) are stated and proved in terms of real-valued functions. It can be seen from Lemma 1, or otherwise directly from the proofs, that these results are in fact valid for $\mathbf{S}$-measurable functions in general.

LEMMA 2. *A function f on X into R is $\mathbf{S}$-measurable if and only if f is an $\mathbf{S}_f$-measurable function.*

PROOF. Since $\mathbf{S}_f \subseteq \mathbf{S}$ in any case, $f^{-1}(\mathbf{R}) \subseteq \mathbf{S}_f$ implies $f^{-1}(\mathbf{R}) \subseteq \mathbf{S}$. Conversely, if f is $\mathbf{S}$-measurable, then $A \in f^{-1}(\mathbf{R})$ implies $A \in \mathbf{S}_f$, by the definition of $\mathbf{S}_f$, so that $f^{-1}(\mathbf{R}) \subseteq \mathbf{S}_f$. This completes the proof.

THEOREM 1. *A necessary and sufficient condition that every subfield of $\mathbf{S}$ be inducible by a statistic is that X be a countable set.*

PROOF. Suppose first that X is countable, and let there be given a field $\mathbf{S}_0 \subseteq \mathbf{S}$. For each $x \in X$ let E_x be the intersection of all sets $A \subseteq X$ such that $x \in A$ and $A \in \mathbf{S}_0$. Let $D_1, D_2, \cdots$, be an enumeration of the sets E_x such that $D_i \cap D_j$ is empty for $i \neq j$ and $\bigcup_i D_i = X$. Define $f(x) = i$ for $x \in D_i$ $(i = 1, 2, \cdots)$. We shall show that $\mathbf{S}_0 = \mathbf{S}_f$.

Since X is separable in the discrete topology, the intersection of any collection of subsets of X equals the intersection of a countable subcollection. Hence $D_i \in \mathbf{S}_0$ for each i. Since $\mathbf{S}_f = \{f^{-1}(N): N \subseteq I\}$ where I is the set of positive integers, and $f^{-1}(N) = \bigcup_{i \in N} D_i$, it follows that $A \in \mathbf{S}_f$ implies $A \in \mathbf{S}_0$. To prove the converse, choose and fix an $A \in \mathbf{S}_0$. Since $x \in E_x$ for all x, we have $A \subseteq \bigcup_{x \in A} E_x$; on the other hand, $E_x \subseteq A$ for each $x \in A$, so that $\bigcup_{x \in A} E_x \subseteq A$; hence $A = \bigcup_{x \in A} E_x = \bigcup_{i \in N} D_i = f^{-1}(N)$ for some $N \subseteq I$. Thus $A \in \mathbf{S}_0$ implies $A \in \mathbf{S}_f$. Hence $\mathbf{S}_f = \mathbf{S}_0$. Since $\mathbf{S}_0$ is arbitrary, the first part of the theorem is proved.

To prove the second part suppose that X is an uncountable set. Let $\mathbf{S}^*$ be the class of all sets A such that one of the sets A and $X - A$ is countable. Then $\mathbf{S}^*$ is a subfield of $\mathbf{S}$ such that for each $x \in X$ the set $\{x\}$ belongs to $\mathbf{S}^*$. Moreover, it can be shown that $\mathbf{S}^* \neq \mathbf{S}$, that is to say, there exists at least one $A \in \mathbf{S}$ such that neither A nor $X - A$ is countable. It follows from Lemma 1 of [2] that there exists no f such that $\mathbf{S}^* = \mathbf{S}_f$. This completes the proof.

A subfield $\mathbf{S}_0$ is *separable* if there exists a countable class $\mathbf{C}$ of subsets of X such that $\mathbf{S}_0$ is the field generated by $\mathbf{C}$. While $\mathbf{S}$ itself is separable, a given subfield $\mathbf{S}_0$ may or may not be separable.[3] However, we have:

LEMMA 3. *Corresponding to any subfield $\mathbf{S}_0$ there exists a separable subfield $\mathbf{S}^*$ such that $\mathbf{S}^* = \mathbf{S}_0$ $[\mathbf{S}, \mu]$.*

PROOF. For the purposes of this proof only, for any two sets A and B in $\mathbf{S}$

---

[3] The writer is indebted to Professor A. Dvoretzky for this remark.

write $A \backsim B$ if and only if $(A - B) \cup (B - A)$ is of μ-measure zero. Let $\{\mathfrak{C}_\theta: \theta \,\varepsilon\, \Omega\}$ be the set of equivalence classes generated by the relation $\backsim$, where Ω is an index set of points θ. For any θ and δ in Ω define $\rho(\theta, \delta) = \mu(A - B) + \mu(B - A)$, where A and B are sets in $\mathfrak{C}_\theta$ and $\mathfrak{C}_\delta$ respectively. Since $\mathbf{S}$ is separable, Ω is a separable metric space under the metric ρ ([5], p. 168).

Let Ω_0 be the set of all θ such that $\mathfrak{C}_\theta$ contains at least one $\mathbf{S}_0$-measurable set. Then Ω_0 is a nonempty subset of Ω and therefore separable. Let Ω^* be a countable subset of Ω_0 which is dense in Ω_0. For each θ in Ω^* let A_θ be an $\mathbf{S}_0$-measurable set in $\mathfrak{C}_\theta$, and let $\mathbf{S}^*$ be the field generated by the class $\{A_\theta: \theta \,\varepsilon\, \Omega^*\}$. It is clear that $\mathbf{S}^*$ is a separable field, and that $\mathbf{S}^* \subseteq \mathbf{S}_0$. We proceed to show that $\mathbf{S}_0 \subseteq \mathbf{S}^*\,[\mathbf{S}, \mu]$.

Choose and fix an $A \,\varepsilon\, \mathbf{S}_0$. By the definition of Ω_0, there exists a $\theta \,\varepsilon\, \Omega_0$ such that $A \,\varepsilon\, \mathfrak{C}_\theta$. Since Ω^* is dense in Ω_0, there exists a sequence $\{\theta_n\}$ in Ω^* such that $\lim_{n\to\infty} \rho(\theta_n, \theta) = 0$. Letting f_0 denote the characteristic function of the set A, it follows from the definition of $\mathbf{S}^*$ that there exists a sequence $f_1, f_2, \cdots$ of $\mathbf{S}^*$-measurable characteristic functions such that $\lim_{n\to\infty} f_n = f_0$ in measure. Hence there exists a subsequence of $\{f_n\}$, say $\{g_n\}$, such that, except on an $\mathbf{S}$-μ-null set, $\lim_{n\to\infty} g_n(x) = f_0(x)$([5], p. 93). Let B be the set of all x such that $\lim_{n\to\infty} g_n(x) = 1$. Then B is $\mathbf{S}^*$-measurable, and $A \backsim B$. Since A is arbitrary, we conclude that $\mathbf{S}_0 \subseteq \mathbf{S}^*\,[\mathbf{S}, \mu]$. This completes the proof.

LEMMA 4. *If $\mathbf{S}^*$ is a separable subfield, there exists an $\mathbf{S}$-measurable function f on X into R such that $f^{-1}(\mathbf{R}) = \mathbf{S}^*$.*

PROOF. Suppose that $\mathbf{S}^*$ is generated by $\mathbf{C} = \{A_1, A_2, \cdots\}$. Let ϕ_i be the characteristic function of A_i, and define $\psi(x) = (\phi_1(x), \phi_2(x), \cdots)$. Suppose that ψ takes values in the space R^n of points $r_{(n)} = (r_1, r_2, \cdots)$ for $1 \leq n \leq \infty$. Since $A_i = \{x: \phi_i(x) = 1\} = \psi^{-1}(B_i)$, where $B_i = \{r_{(n)}: r_i = 1\}$, we have $A_i \,\varepsilon\, \psi^{-1}(\mathbf{R}^n)$ for each i; hence $\mathbf{S}^* \subseteq \psi^{-1}(\mathbf{R}^n)$. On the other hand, since each ϕ_i is an $\mathbf{S}^*$-measurable function, ψ is $\mathbf{S}^*$-measurable also, so that $\psi^{-1}(\mathbf{R}^n) \subseteq \mathbf{S}^*$. Thus $\mathbf{S}^* = \psi^{-1}(\mathbf{R}^n)$; the lemma as stated now follows from Lemma 1 by taking $f = \alpha_n \psi$.

LEMMA 5. *If f is an $\mathbf{S}$-measurable function on X into R, then $f^{-1}(\mathbf{R}) = \mathbf{S}_f\,[\mathbf{S}, \mu]$.*

PROOF. According to Lemma 2, $f^{-1}(\mathbf{R}) \subseteq \mathbf{S}_f$. We have therefore to show that $\mathbf{S}_f \subseteq f^{-1}(\mathbf{R})\,[\mathbf{S}, \mu]$.

We recall that we have assumed $X \subseteq R^m$, $X \,\varepsilon\, \mathbf{R}^m$, and $\mathbf{S} = \{X \cap A: A \,\varepsilon\, \mathbf{R}^m\}$. Let α_m be the function described in Lemma 1, and write $\alpha_m(X) = Y$, $\alpha_m(\mathbf{S}) = \mathbf{T}$, $g(y) = f(\alpha_m^{-1}(y))$ for $y \,\varepsilon\, Y$. Then Y is a Borel set of the real line, $\mathbf{T}$ is the class of Borel sets of Y, g is a $\mathbf{T}$-measurable function on Y into R, and $f^{-1}(\mathbf{R}) = \alpha_m^{-1}(g^{-1}(\mathbf{R}))$, $\mathbf{S}_f = \alpha_m^{-1}(\mathbf{T}_g)$. Define $\nu(C) = \mu(\alpha_m^{-1}(C))$ for $C \,\varepsilon\, \mathbf{T}$. It is then easily seen that the desired conclusion is equivalent to $\mathbf{T}_g \subseteq g^{-1}(\mathbf{R})\,[\mathbf{T}, \nu]$.

Choose and fix a set $A \,\varepsilon\, \mathbf{T}_g$. By definition of $\mathbf{T}_g$, there exists a set $B \subseteq R$ such that $g^{-1}(B) = A$. Now, since A is a Borel set, and g is a Borel measurable function, it follows from Lusin's theorem ([6], p. 72) that for each $k = 1, 2 \cdots$ there exists a set $A_k \,\varepsilon\, \mathbf{T}$ such that $A_k \subseteq A$, $\nu(A - A_k) < 1/k$, and $g(A_k) \,\varepsilon\, \mathbf{R}$.

Let $A_0 = \bigcup_k A_k$. Then $A_0 \varepsilon \mathbf{T}$, $A_0 \subseteq A$, $\nu(A - A_0) = 0$, and $g(A_0) = \bigcup_k g(A_k) = B_0$ (say) is a Borel set. Now, $g^{-1}(B_0) = g^{-1}(g(A_0)) \supseteq A_0$. Also, $B_0 = g(A_0) \subseteq g(A) = B$, so that $g^{-1}(B_0) \subseteq g^{-1}(B) = A$. Hence $C = g^{-1}(B_0)$ is a set such that $A_0 \subseteq C \subseteq A$, so that $\nu(A - C) = 0$; since C is a set in $g^{-1}(\mathbf{R})$, and since $A \ \varepsilon \ \mathbf{T}_g$ in this argument is arbitrary, it follows that $\mathbf{T}_g \subseteq g^{-1}(\mathbf{R})$ $[\mathbf{T}, \nu]$. This completes the proof.

REMARK 1. The preceding argument shows that μ is a perfect measure, i.e., for each real-valued $\mathbf{S}$-measurable function f, corresponding to each set A in $\mathbf{S}_f$ there exists a B in $f^{-1}(\mathbf{R})$ such that $B \subseteq A$ and $\mu(A - B) = 0$. (Cf. [9], p. 18; also pp. 248–251.) Perfection is a little stronger than the property stated in Lemma 5. The fact that μ is perfect can be deduced, alternatively, from Theorem 1 of [9], p. 18, since $(X, \mathbf{S}, \mu)$ is a euclidean space.

REMARK 2. If X is an uncountable set, the "exact" form of Lemma 5 is false, that is to say, there do exist $\mathbf{S}$-measurable functions f for which $f^{-1}(\mathbf{R}) \neq \mathbf{S}_f$. This follows easily from the theory of analytic sets [7].

As an immediate consequence of Lemmas 3, 4, and 5 we have:

THEOREM 2. *Corresponding to any subfield* $\mathbf{S}_0$ *there exists an* $\mathbf{S}$*-measurable function f on X into R such that* $\mathbf{S}_f = \mathbf{S}_0$ $[\mathbf{S}, \mu]$.

The remainder of this section is devoted to showing that, for $\mathbf{S}$-measurable functions, contraction coincides with functional contraction (Theorem 3).

LEMMA 6. *If f is an* $\mathbf{S}$*-measurable function on X into R, and* $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$, *then* $f \subseteq g$ $[\mathbf{S}, \mu]$.

PROOF. Since f is $\mathbf{S}_f$-measurable (cf. Lemma 2), the hypothesis $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$ yields the existence of an $\mathbf{S}_g$-measurable function, h say, such that the set $\{x: f(x) \neq h(x)\}$ is $\mathbf{S}$-μ-null. Denote this last set by N, and let $g(X - N) = A$.

Since h is $\mathbf{S}_g$-measurable, it depends on x only through g (cf. Lemma 3.2 of [1]), say $h(x) = k(g(x))$ for all x. Define $k^* = k$ on A and $= \alpha$ on $g(X) - A$, where α is a point in $f(X)$. Then k^* is a function on the range of g into that of f such that $\{x: f(x) \neq k^*(g(x))\}$ is a subset of N; this completes the proof.

Let $\bar{\mathbf{S}}$ be the class of all sets of the form $A \cup C$ where A is $\mathbf{S}$-measurable and C is a subset of an $\mathbf{S}$-μ-null set, and define $\bar{\mu}(A \cup C) = \mu(A)$. Then $\bar{\mathbf{S}}$ is a field containing $\mathbf{S}$, $\bar{\mu}$ is a probability measure on $\bar{\mathbf{S}}$, and $\bar{\mu}(A) = \mu(A)$ for $A \ \varepsilon \ \mathbf{S}$. For any statistic f, $\bar{\mathbf{S}}_f$ is defined, as usual, as the class of all $\bar{\mathbf{S}}$-measurable sets of the form $f^{-1}(B)$. (Note. In general, $\bar{\mathbf{S}}_f$ is different from $\overline{(\mathbf{S}_f)}$.)

LEMMA 7. *If* $f \subseteq g$ $[\bar{\mathbf{S}}, \bar{\mu}]$, *then* $\bar{\mathbf{S}}_f \subseteq \bar{\mathbf{S}}_g$ $[\bar{\mathbf{S}}, \bar{\mu}]$.

PROOF. By hypothesis, there exists a function h on the range of g into that of f, and an $\bar{\mathbf{S}}$-$\bar{\mu}$-null set N such that $f(x) = h(g(x))$ on $X - N$. Choose and fix a set in $\bar{\mathbf{S}}_f$, say $A = f^{-1}(B)$. Define $A^* = g^{-1}(C)$, where $C = h^{-1}(B)$.

Write $N^* = \{x : f(x) \neq h(g(x))\}$. Then $N^* \subseteq N$, so that N^* is an $\bar{\mathbf{S}}$-$\bar{\mu}$-null set. We have $A^* \cap (X - N^*) = \{x: g \ \varepsilon \ h^{-1}(B), f = hg\} = \{x: hg \ \varepsilon \ B, f = hg\} = \{x: f \ \varepsilon \ B, f = hg\} = A \cap (X - N^*)$. Hence $A^* - A \subseteq N^*$ and $A - A^* \subseteq N^*$. Since A is $\bar{\mathbf{S}}$-measurable and $\bar{\mu}$ is complete on $\bar{\mathbf{S}}$, it follows that A^* is $\bar{\mathbf{S}}$-measurable (and therefore in $\bar{\mathbf{S}}_g$) and that A^* differs from A by a set of $\bar{\mu}$-measure zero. Since $A \ \varepsilon \ \bar{\mathbf{S}}_f$ is arbitrary, the lemma is proved.

LEMMA 8. *If g is an* $\mathbf{S}$*-measurable function on X into R, then* $\mathbf{S}_g = \bar{\mathbf{S}}_g$ $[\bar{\mathbf{S}}, \bar{\mu}]$.

PROOF. Since $\mathbf{S} \subseteq \bar{\mathbf{S}}$, we have $\mathbf{S}_g \subseteq \bar{\mathbf{S}}_g$; and since g is $\mathbf{S}$-measurable, $g^{-1}(\mathbf{R}) \subseteq \mathbf{S}_g$ by Lemma 2. Thus $g^{-1}(\mathbf{R}) \subseteq \mathbf{S}_g \subseteq \bar{\mathbf{S}}_g$. The desired conclusion can now be established by showing that $\bar{\mathbf{S}}_g \subseteq g^{-1}(\mathbf{R})$ $[\bar{\mathbf{S}}, \bar{\mu}]$. The demonstration of this last relation is essentially the same as the proof of the nontrivial part of Lemma 5, and so is omitted.

LEMMA 9. *If g is an $\mathbf{S}$-measurable function on X into R, and $f \subseteq g$ $[\mathbf{S}, \mu]$, then $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$.*

PROOF.
$$\begin{aligned} f \subseteq g\,[\mathbf{S}, \mu] &\leftrightarrow f \subseteq g\,[\bar{\mathbf{S}}, \bar{\mu}] \\ &\rightarrow \bar{\mathbf{S}}_f \subseteq \bar{\mathbf{S}}_g\,[\bar{\mathbf{S}}, \bar{\mu}] \qquad \text{(Lemma 7)} \\ &\leftrightarrow \bar{\mathbf{S}}_f \subseteq \mathbf{S}_g\,[\bar{\mathbf{S}}, \bar{\mu}] \qquad \text{(Lemma 8)} \\ &\rightarrow \mathbf{S}_f \subseteq \mathbf{S}_g\,[\bar{\mathbf{S}}, \bar{\mu}] \\ &\leftrightarrow \mathbf{S}_f \subseteq \mathbf{S}_g\,[\mathbf{S}, \mu]. \end{aligned}$$

THEOREM 3. *Let f and g be $\mathbf{S}$-measurable functions on X into R. Then $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$ if an only if $f \subseteq g$ $[\mathbf{S}, \mu]$.*

The proof is immediate from Lemmas 6 and 9.

It can be shown by the methods used in this section that Theorems 1, 2, and 3 are valid for any probability space $(X, \mathbf{S}, \mu)$ which satisfies the following conditions: (i) for each x in X, $\{x\}$ is $\mathbf{S}$-measurable, (ii) $\mathbf{S}$ is separable, and (iii) μ is perfect. However, such a space can differ but little from a euclidean sample space.

3. Applications to the theory of sufficiency. We suppose now that there is given a euclidean sample space $(X, \mathbf{S})$, as before, and a dominated set P of probability measures on $\mathbf{S}$. Definitions of the technical terms used here without explanation are given in the first part of [1]. The conclusions of this section are relevant to problem 3 of [1], p. 441.

Let μ be an arbitrary but fixed probability measure, not necessarily in P, such that for each $\mathbf{S}$-measurable set A, $\mu(A) = 0$ if and only if $p(A) = 0$ for each p in P. The existence of such a μ is assured by Lemma 7 of [8].

COROLLARY 1. *There exists a function f on X into R such that:*

(a) f is $\mathbf{S}$-measurable,

(b) $\mathbf{S}_f$ is a necessary and sufficient subfield,

(c) f is a necessary and sufficient statistic.

PROOF. Since P is dominated, it follows from Theorem 6.2 of [1] that there exists a subfield $\mathbf{S}_0$ (say) which is necessary and sufficient. Let f be a function on X into R such that (a) holds, and such that $\mathbf{S}_f = \mathbf{S}_0$ $[\mathbf{S}, \mu]$; such an f exists, by Theorem 2. Property (b) is immediate (cf. Corollary 6.2 (iii) of [1]), and it remains to verify (c). Since $\mathbf{S}_f$ is sufficient ($\equiv f$ is sufficient) by (b), we have only to show that f is a necessary statistic. Let g be any sufficient statistic. Then $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$, since $\mathbf{S}_g$ is sufficient by hypothesis and $\mathbf{S}_f$ is necessary by (b). Hence $f \subseteq g$ $[\mathbf{S}, \mu]$, by (a) and Lemma 6. This completes the proof.

REMARK. It is evident from Lemma 5 that Corollary 1 remains valid if $\mathbf{S}_f$ is replaced by $f^{-1}(\mathbf{R})$ in (b). It can be shown that this modified version of Corollary 1 is valid not only in the present case but in any framework $(X, \mathbf{S})$, P

provided that P is a separable metric space under the metric $\delta(p, q) = \sup_{A \varepsilon \mathbf{S}} |p(A) - q(A)|$.

COROLLARY 2. *If g is a necessary and sufficient statistic, then $\mathbf{S}_g$ is a necessary and sufficient subfield.*

PROOF. We have only to show that if g is a necessary statistic, then $\mathbf{S}_g$ is a necessary subfield. Let f be a function on X into R such that conditions (a), (b), and (c) of Corollary 1 are satisfied. Since f is sufficient and g is necessary, we have $g \subseteq f$ $[\mathbf{S}, \mu]$. Hence $\mathbf{S}_g \subseteq \mathbf{S}_f$ $[\mathbf{S}, \mu]$ by Lemma 9. Since $\mathbf{S}_f$ is necessary, it follows that $\mathbf{S}_g$ is necessary, and the proof is complete.

It should be stated here that the converse of Corollary 2 is false (cf. Example 2 in Section 4), and also that the corollary itself is false in the general case (cf. [2]).

COROLLARY 3. *Let g be an $\mathbf{S}$-measurable function on X into R. Then g is a necessary and sufficient statistic if and only if $g^{-1}(\mathbf{R})$ is a necessary and sufficient subfield.*

PROOF. In view of Corollary 2 and Lemma 5, we have only to show that if $\mathbf{S}_g$ is a necessary subfield, then g is a necessary statistic; since g is $\mathbf{S}$-measurable, the desired result follows from Lemma 6 by the argument used in establishing part (c) of Corollary 1.

4. Two examples. In both examples, $X = U \times V$ is the set of all points $x = (u, v)$ with $-\infty < u < \infty$, $-\infty < v < \infty$; $\mathbf{S}$ is the field of Borel sets of X; $P = \{p_\theta : -\infty < \theta < \infty\}$, where p_θ is the measure on $\mathbf{S}$ corresponding to u and v being independent normally distributed random variables, with means θ and 0 respectively and variances 1; and $\mu = p_{\theta=0}$. Let U and V denote, respectively, the coordinate axes $v = 0$ and $u = 0$. Let $\mathbf{U}$ and $\mathbf{V}$ denote, respectively, the Borel sets of U and V.

The first example shows that the following propositions are false:

(i) If $f \subseteq g$ $[\mathbf{S}, \mu]$, then $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$. (Cf., however, Lemmas 7 and 9.)

(ii) If f is sufficient, and f is a functional contraction of g (that is, $f \subseteq g$ $[\mathbf{S}, \mu]$), then g is sufficient. (Cf. Theorem 6.4 of [1].)

EXAMPLE 1. Let $f(u, v) \equiv u$. To define g, let $N \subseteq V$ be a set such that N has linear measure zero but is not in $\mathbf{V}$. Let $g(u, v) = u$ for $v \varepsilon V - N$, and $g(u, v) = 0$ for $v \varepsilon N$. Then $f \subseteq g$ $[\mathbf{S}, \mu]$, and also $g \subseteq f$ $[\mathbf{S}, \mu]$ so that f and g are functionally equivalent. However, it is easily seen from Fubini's theorem ([6], p. 83) that $\mathbf{S}_f = f^{-1}(\mathbf{U})$ while $\mathbf{S}_g$ contains only X and the empty set.

The second example shows that the following propositions are false:

(iii) If $\mathbf{S}_f \subseteq \mathbf{S}_g$ $[\mathbf{S}, \mu]$, then $f \subseteq g$ $[\mathbf{S}, \mu]$. (Cf., however, Lemma 6.)

(iv) If S_f is a necessary and sufficient subfield, then f is a necessary and sufficient statistic. (Cf., however, Corollary 3 together with Lemma 5.)

EXAMPLE 2. Define $g(u, v) \equiv u$. To define f, let $M \subseteq V$ be a set which is not measurable with respect to linear measure on V, and let $f(u, v) = (u, 1)$ for $v \varepsilon M$ and $f(u, v) = (u, 2)$ for $v \varepsilon V - M$. We shall show that $\mathbf{S}_f = \mathbf{S}_g$, so that f and g are equivalent, but that f is not a functional contraction of g.

Let $U_1 = \{x : v = 1\}$, $U_2 = \{x : v = 2\}$. Since g is *exactly* a function of f,

$\mathbf{S}_g \subseteq \mathbf{S}_f$. To prove the converse, consider a fixed $C \varepsilon \mathbf{S}_f$. There exists a set $B_1 \subseteq U_1$ and a $B_2 \subseteq U_2$ such that $C = f^{-1}(B_1 \cup B_2)$. Let the perpendicular projections of B_1, B_2 on U be A_1, A_2, respectively. Then, by the definition of f, $C = E \cup F \cup G$, where $E = (A_1 \cap A_2) \times V$, $F = (A_1 - A_2) \times M$, and $G = (A_2 - A_1) \times (V - M)$. Since C is a Borel set while M and $V - M$ are not, it follows ([6], p. 83) that F and G must be empty. Hence $A_1 = A_2 = A$ say, and $C = A \times V = g^{-1}(A)$. It now follows ([6], p. 83) that A is in $\mathbf{U}$, so that $g^{-1}(A) = C$ is in $g^{-1}(\mathbf{U})$. Since C is arbitrary, we have $\mathbf{S}_f \subseteq g^{-1}(\mathbf{U})$; but $g^{-1}(\mathbf{U}) = \mathbf{S}_g$, so that $\mathbf{S}_f \subseteq \mathbf{S}_g$. Thus $\mathbf{S}_f = \mathbf{S}_g$.

To show that f is not a functional contraction of g, suppose to the contrary that $f \subseteq g$ $[\mathbf{S}, \mu]$. Then $f \subseteq g$ $[\bar{\mathbf{S}}, \bar{\lambda}]$, where $\bar{\mathbf{S}}$ denotes the Lebesgue measurable sets of X and $\bar{\lambda}$ is (planar) Lebesgue measure on $\bar{\mathbf{S}}$. In other words, there exists a function h on U into $U_1 \cup U_2$ and an $\bar{\mathbf{S}}$-$\bar{\lambda}$-null set N such that $f(x) = h(g(x))$ on $X - N$. Write $h^{-1}(U_1) = I$, $U \times M = J$, and $I \times V = K$. Then $J = f^{-1}(U_1)$ and $K = g^{-1}(h^{-1}(U_1))$, and it follows exactly as in the proof of Lemma 7 that the sets $J - K$ and $K - J$ are $\bar{\mathbf{S}}$-$\bar{\lambda}$-null. Hence $L = (J - K) \cup (K - J)$ is $\bar{\mathbf{S}}$-$\bar{\lambda}$-null.

For each $u \varepsilon U$, let E_u be the set of all $v \varepsilon V$ such that $(u, v) \varepsilon L$. Let $\bar{\lambda}_u$, $\bar{\lambda}_v$ denote linear measure on U, V, respectively. It follows from Fubini's theorem ([6], p. 81) that there exists a $C \subseteq U$ with $\bar{\lambda}_u(C) = 0$ such that, for each $u \varepsilon U - C$, the set E_u is $\bar{\lambda}_v$-measurable (and of $\bar{\lambda}_v$-measure zero). Since $u \varepsilon I$ implies $E_u = V - M$, and $u \varepsilon U - I$ implies $E_u = M$, and since at least one of the sets $I - C$, $U - I - C$ must be nonempty (because $\bar{\lambda}_u(C) = 0$), it follows that M is $\bar{\lambda}_v$-measurable, and this is a contradiction.

It can be shown by a slight elaboration of the preceding argument that in Example 2 we have $\bar{\mathbf{S}}_f \subseteq \bar{\mathbf{S}}_g$ $[\bar{\mathbf{S}}, \bar{\mu}]$, but not $f \subseteq g$ $[\bar{\mathbf{S}}, \bar{\mu}]$. This, together with Lemma 7, shows that by completing a given probability space $(X, \mathbf{S}, \mu)$ to $(X, \bar{\mathbf{S}}, \bar{\mu})$ the inconsistency between contraction and functional contraction is reduced but not eliminated entirely.

REFERENCES

[1] R. R. Bahadur, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[2] R. R. Bahadur and E. L. Lehmann, "Two comments on sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 139–141.

[3] E. L. Lehmann and H. Scheffé, "Completeness, similar regions, and unbiased estimation. Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[4] L. J. Savage, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954, p. 112.

[5] P. R. Halmos, *Measure Theory*, Van Nostrand Company, Inc., New York, 1950.

[6] S. Saks, *Theory of the Integral*, Second revised edition, Hafner Publishing Company, New York, 1947.

[7] C. Kuratowski, *Topologie I*, Deuxième Édition, Monografie Matematyczne, Warszawa, 1948.

[8] P. R. Halmos and L. J. Savage, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 225–241.

[9] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions of Sums of Independent Random Variables* (translated from the Russian by K. L. Chung), Addison-Wesley Publishing Co., Cambridge, Mass., 1954.

ESTIMATION OF PARAMETERS OF TRUNCATED OR CENSORED EXPONENTIAL DISTRIBUTIONS

By Walter L. Deemer, Jr. and David F. Votaw, Jr.

United States Air Force and Yale University

1. Summary. This paper gives maximum likelihood estimators of parameters of truncated and censored exponential distributions, asymptotic variances of the estimators, and asymptotic confidence intervals for the parameters.

Applications to bombing accuracy studies and to life testing are pointed out. As regards bombing accuracy the parameter estimated is the reciprocal of the variance in a normal bivariate distribution having circular symmetry. The reciprocal is estimated because there is no maximum likelihood estimator of the variance and any estimator of the variance is badly biased (see Section 2).

Results of a synthetic sampling experiment are given to provide information on rapidity of convergence of the distributions of the estimators to their asymptotic distributions.

2. Introduction. In bombing accuracy studies and in other aiming accuracy studies, the assumption is often made that aiming errors (range and deflection errors in bombing; azimuth and elevation errors in gunnery) have a bivariate normal distribution with mean at the aiming point, zero correlation and equal variances.

Under these assumptions the radial error, or distance from the aiming point to the point of impact, is a chance quantity say R with probability density function

$$k(r) = r\sigma^{-2} \exp [-r^2/(2\sigma^2)], \qquad 0 < r < \infty. \tag{2.1}$$

Let $\frac{1}{2}R^2 = Z$, say, and denote σ^{-2} by c. The density, say $h(z)$, of Z is

$$h(z) = ce^{-cz}, \qquad 0 < z < \infty ; c > 0; \tag{2.2}$$

thus Z has an exponential distribution.

In some situations values of Z greater than a fixed value cannot be observed. For example, in gun camera missions the view angle of the camera defines the maximum observable R (and thus the maximum observable Z). An example arises in life testing from an exponential distribution when the time of testing is fixed in advance (see [3], pp. 4–9). (Cases in which the time of testing is determined by a sample are treated in [1], [3], [4], and [6], p. 416.)

Before proceeding with the estimation in truncated and censored cases let us consider estimation[1] of c in (2.2) on the basis of a sample $Z_1, Z_2, \cdots, Z_N$

Received November 4, 1953; revised November 20, 1954.

[1] For estimation of $\sigma(= c^{-1/2})$ when the observations are grouped see [5].

of values of Z. The likelihood function, $L(c)$, of c is

$$L(c) = [ce^{-c\bar{z}}]^N \tag{2.3}$$

where $\bar{z}$ is the sample mean. The value, say $\hat{c}$, of c for which $L(c)$ assumes its maximum value is

$$\hat{c} = (\bar{z})^{-1}, \text{ the maximum likelihood estimator of } c. \tag{2.4}$$

The estimator $\hat{c}$ has a finite mean if $N \geqq 2$, and a finite variance if $N \geqq 3$.

It is well known that $2Nc/\hat{c}$ has a chi-square distribution with $2N$ degrees of freedom. Equation (2.4) is equivalent to the well-known result that the maximum likelihood estimator, say $\hat{\sigma}^2$, of σ^2 is

$$\hat{\sigma}^2 = \sum_{i=1}^{N} r_i^2/2N. \tag{2.5}$$

The asymptotic variance of $(N)^{1/2}\ (\hat{c} - c)$ is

$$[-E(\partial^2 \log h(z)/\partial c^2)]^{-1}.$$

From (2.2) we have that this equals c^2; therefore, for large N

$$\text{Variance } [(N)^{1/2}\ (\hat{c} - c)] = c^2. \tag{2.6}$$

Derivations of the asymptotic variance of a maximum likelihood estimator are given in [6], pp. 208–212, and [7], pp. 136–139.

When the distribution is truncated or censored, we shall replace Z by X and denote by x_0 the maximum value of X that can be observed. It is assumed that x_0 is known in advance. The two cases will now be described.

Case A (*Censored*[2] *Distribution*). Here the number of observations greater than x_0 is known. When $Z \leqq x_0$, $X = Z$; when $Z > x_0$, the only information obtained about X is simply that $X > x_0$. X can be regarded as having a density, say $g(x)$, when $X \leqq x_0$; thus

$$\begin{aligned} g(x) &= ce^{-cx}, \qquad 0 < x \leqq x_0, \\ \Pr(X > x_0) &= e^{-cx_0}. \end{aligned} \tag{2.7}$$

Case B (*Truncated Distribution*). Here the number of observations greater than x_0 is unknown. X has a density, say $f(x)$, which is the conditional density of Z given that $Z \leqq x_0$; thus

$$f(x) = ce^{-cx}\ (1 - e^{-cx_0})^{-1}, \qquad 0 < x \leqq x_0. \tag{2.8}$$

The maximum likelihood estimator of c will be derived for Case A and for Case B. It is noteworthy that in each case no maximum likelihood estimator of $\sigma^2\ (= c^{-1})$ exists and the bias of any estimator of σ^2 tends to $-\infty$ as σ^2 tends to $+\infty$. For this reason the quantity c instead of c^{-1} is chosen as the parameter to be estimated.

[2] For further discussion of censored and truncated distributions see [2], p. 144.

3. Maximum-likelihood estimators. For Case A let n be the number of observations of X such that $X \leqq x_0$ and let m be the number of observations such that $X > x_0$. Let $N = m + n$. The likelihood function, say $L_A(c)$, of c is (see (2.7)),

$$L_A(c) = \begin{cases} N!\,[m!\,n!]^{-1} c^n \exp\left[-c \sum_1^n x_i - mcx_0\right], & n > 0 \\ e^{-Ncx_0}, & n = 0. \end{cases} \tag{3.1}$$

(It should be noted that this is the likelihood function of a chance quantity having the density given in (2.7) and a probability e^{-cx_0} of taking the value x_0. Halperin [3], pp. 4–9, has proved that the maximum likelihood estimator of c in this mixed continuous discrete case has the properties of consistency, asymptotic normality, and minimum asymptotic variance.)

The maximum-likelihood estimator, say $\hat{c}_A$, of c is

$$\hat{c}_A = n\left[mx_0 + \sum_1^n x_i\right]^{-1}. \tag{3.2}$$

$\hat{c}_A$ has a finite mean if $N \geqq 2$ and a finite variance if $N \geqq 3$.

For Case B let the sample be $X_1, \cdots, X_{n'}$. The likelihood function, say $L_B(c)$, is (see (2.8)),

$$\begin{aligned} L_B(c) &= c^{n'}(1 - e^{-cx_0})^{-n'} \exp\left[-c \sum_1^{n'} x_i\right] \\ &= [ce^{-c\bar{x}}(1 - e^{-cx_0})^{-1}]^{n'}, \end{aligned} \tag{3.3}$$

where $\bar{x}$ is the sample mean. It follows that

$$\partial \log L_B(c)/\partial c = n'[c^{-1} - x_0 e^{-cx_0}(1 - e^{-cx_0})^{-1} - \bar{x}]. \tag{3.4}$$

It can be shown that the function $c^{-1} - x_0 e^{-cx_0}(1 - e^{-cx_0})^{-1}$ is monotonic decreasing in c; as c tends to 0 the function tends to $\frac{1}{2}x_0$, and as c tends to infinity the function tends to 0. When $0 < \bar{x} < \frac{1}{2}x_0$, there exists a solution, say c', of the equation formed by setting $\partial \log L_B(c)/\partial c$ equal to 0 (see (3.4)). Clearly c' is the maximum likelihood estimator of c when $0 < \bar{x} < \frac{1}{2}x_0$. When $\bar{x} \geqq \frac{1}{2}x_0$, the function $L_B(c)$ assumes its maximum value for $c = 0$. The maximum likelihood estimator, say $\hat{c}_B$, of c can be described as follows:

$$\hat{c}_B = \begin{cases} c', & \text{when } 0 < \bar{x} < \frac{1}{2}x_0 \\ 0, & \text{when } \bar{x} \geqq \frac{1}{2}x_0. \end{cases} \tag{3.5}$$

A table of $\bar{x}/x_0$ as a function of $c'x_0$ is given in Table 1.

The estimator $\hat{c}_B$ is less than $n'(\sum_1^{n'} x_i)^{-1}$, which is the estimator $\hat{c}$ when $n' = N$ (see (2.4)). This follows from the fact that when $n' = N$, $L_B(c) = L(c)(1 - e^{-cx_0})^{-n'}$ (see (2.3), (3.3)). The estimator $\hat{c}_B$, therefore, has finite mean for $n' \geqq 2$ and finite variance for $n' \geqq 3$.

TABLE 1

$$\frac{\bar{x}}{x_0} = \frac{1}{c'x_0} - \frac{1}{e^{c'x_0} - 1}$$

$c'x_0$	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.		.4916	.4832	.4750	.4668	.4584	.4504	.4422	.4340	.4260
1.	.4180	.4102	.4024	.3946	.3870	.3794	.3720	.3648	.3576	.3504
2.	.3434	.3366	.3300	.3234	.3168	.3106	.3044	.2984	.2924	.2866
3.	.2810	.2754	.2700	.2648	.2596	.2546	.2496	.2450	.2402	.2358
4.	.2314	.2270	.2228	.2188	.2148	.2110	.2072	.2036	.2000	.1966
5.	.1932	.1900	.1868	.1836	.1806	.1778	.1748	.1720	.1694	.1668
6.	.1642	.1616	.1592	.1568	.1546	.1524	.1502	.1480	.1460	.1440
7.	.1420	.1400	.1382	.1364	.1346	.1328	.1310	.1294	.1278	.1262
8.	.1246	.1232	.1216	.1202	.1188	.1174	.1160	.1148	.1134	.1122

4. Asymptotic variances of the estimators. With regard to Case A we have from results of Halperin [3], pp. 4–9, that the asymptotic variance of $(N)^{1/2} \cdot (\hat{c}_A - c)$ is the reciprocal of

$$\int_0^{x_0} \left[\frac{\partial \log g(x)}{\partial c}\right]^2 g(x)\, dx + q\left[\frac{\partial \log q}{\partial c}\right]^2, \tag{4.1}$$

where $q = \Pr(X > x_0) = e^{-cx_0}$ (see (2.7)).

The expression in (4.1) equals

$$c^{-2}(1 - e^{-cx_0}); \tag{4.2}$$

accordingly, for large N

$$\text{Variance } [(N)^{1/2}(\hat{c}_A - c)] = c^2(1 - e^{-cx_0})^{-1}. \tag{4.3}$$

Note that this is always greater than the asymptotic variance of $(N)^{1/2}(\hat{c} - c)$ (see (2.6)).

The asymptotic variance of $(n')^{1/2}(\hat{c}_B - c)$ is the reciprocal of

$$-E(\partial^2 \log f(x)/\partial c^2),$$

where $f(x)$ is given in (2.8). Thus for large n'

$$\text{Variance } [(n')^{1/2}(\hat{c}_B - c)] = [c^{-2} - x_0^2 e^{-cx_0}(1 - e^{-cx_0})^{-2}]^{-1}. \tag{4.4}$$

Having obtained the asymptotic variances of $\hat{c}_A$ and $\hat{c}_B$ let us compare them. The comparison will be made for $n' = N$, which is the most favorable situation for Case B. Let

$$R = \frac{\text{Variance } [(n')^{1/2}(\hat{c}_B - c)]}{\text{Variance } [(n')^{1/2}(\hat{c}_A - c)]}, \quad n' \text{ large.} \tag{4.5}$$

From (4.3) and (4.4) it follows that

$$R = (1 - e^{-cx_0})/[1 - (cx_0)^2 e^{-cx_0}(1 - e^{-cx_0})^{-2}]. \tag{4.6}$$

TABLE 2

Ratio of the Variances of $\hat{c}_B$ and $\hat{c}_A$

$t = cx_0$	$R(t)$	$t = cx_0$	$R(t)$
0.01	1194	1.1	7.02
0.02	594	1.2	6.25
0.04	294	1.3	5.61
0.06	194	1.4	5.07
0.08	144	1.5	4.62
0.1	113	1.6	4.23
0.2	54.6	1.7	3.90
0.3	34.8	1.8	3.61
0.4	24.9	1.9	3.36
0.5	19.1	2.0	3.13
0.6	15.3	3.0	1.89
0.7	12.6	4.0	1.41
0.8	10.7	5.0	1.20
0.9	9.15	10.0	1.00
1.0	7.97		

R can be considered as a function of $cx_0 = t$, say. A table of R as a function of t is given in Table 2. ($R(t) > 1$ for $t > 0$, and $R(t) \to \infty$ as $t \to 0$.)

5. Interval estimation of c. Approximate $100(1 - q)$ per cent confidence limits for c in (2.7) can be obtained by means of the following approximation when the sample size is large:

$$\Pr(-y_q < y < y_q) = q, \tag{5.1}$$

where y_q is the $100(1 - \frac{1}{2}q)$ per cent point of the standard normal distribution and

$$y = N^{1/2}(\hat{c}_A - c)/[c(1 - e^{-cx_0})^{-1/2}]. \tag{5.2}$$

Similarly, when the sample size is large, $100(1 - q)$ per cent confidence limits for c in (2.8) can be obtained by means of (5.1), where

$$y = (n')^{1/2}(\hat{c}_B - c)[c^{-2} - x_0^2e^{-cx_0}(1 - e^{-cx_0})^{-2}]^{1/2}. \tag{5.3}$$

The procedure given in [6], Section 11.7, for constructing confidence limits could be used in the cases discussed above.

6. Synthetic sampling experiment. To throw some light on the rapidity of approach of the distributions of $\hat{c}_A$ and $\hat{c}_B$ to their limiting normal distributions we have carried out a synthetic sampling experiment. With regard to $\hat{c}$ the rapidity of approach can be determined by analytic methods since the exact distribution of $\hat{c}$ is known (see Section 2).

A random sample of 140,000 cases was drawn from a rectangular distribu-

TABLE 3

Synthetic sampling experiment

P is the probability that values of χ^2 as large or larger than that obtained would have been obtained under the null hypothesis.*

Serial No. of set of 20,000 cases	Number of cases per sample	Number of samples	$\hat{c}$		$\hat{c}_A$		$\hat{c}_B$	
			χ^2	P	χ^2	P	χ^2	P
1	100	200	20.2	.38	15.8	.67	18.0	.52
2	100	200	21.6	.31	24.2	.19	29.2	.063
3	100	200	25.6	.14	21.4	.32	14.0	.78
4	200	100	17.6	.55	20.0	.39	20.4	.32
5	200	100	19.2	.44	46.8	.00040	33.2	.023
6	200	100	24.4	.18	26.4	.12	27.2	.10
7	200	100	21.2	.33	12.8	.85	15.2	.71

* Equi-probability intervals (.05) were used throughout; thus there are 19 degrees of freedom.

tion and randomly divided into seven sets of 20,000 cases each. Three of these seven sets were divided into 200 samples of 100 cases each; the other four sets were divided into 100 samples of 200 cases each. The variable with the rectangular distribution was then converted (a) to a variable with density function as given in (2.2) with $c = 1$, and (b) to a variable with density function as given in (2.8) with $c = 1$ and $x_0 = 1$. The variable of (a) was used to calculate $\hat{c}$ for each sample (600 samples of 100 cases each; 400 samples of 200 cases each); this distribution was then censored at $x_0 = 1$ and $\hat{c}_A$ was calculated for each of the 1000 samples. The variable of (b) was used to calculate $\hat{c}_B$ for each of the 1000 samples. The goodness of fit of the limiting normal distributions to the observed distributions of $\hat{c}_A$ and $\hat{c}_B$ was tested by chi-square. The goodness of fit of the exact distribution of $\hat{c}$ to the observed distribution was tested similarly. The chi-square probabilities are given in Table 3. Each of the seven lines of Table 3 represents one of the seven independent sets of 20,000 cases. The three values of the chi-square probability, P, on a given line are not independent because they are based on the same samples.

The results suggest that when cx_0 is as small as 1 and the sample size is as small as 100, the distributions of the estimators are fairly well approximated by the limiting distributions. With less severe limitations (i.e., $cx_0 > 1$) the approximation would be better.

REFERENCES

[1] B. Epstein and M. Sobel, "Some theorems relevant to life testing from an exponential distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 373–381.

[2] A. Hald, *Statistical Theory with Engineering Applications*, John Wiley and Sons, New York, 1952.

[3] Max Halperin, "Empirical study of restriction of range; estimation of parameters in

truncated processes," Report No. 4, USAF School of Aviation Medicine, October 1950.

[4] MAX HALPERIN, "Maximum likelihood estimation in truncated samples," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 226–238.

[5] H. M. HUGHES, "Estimation of the variance of the bivariate normal distribution," *University of California Publications in Statistics*, Vol. 1, No. 4, University of California Press, 1949, pp. 37–52.

[6] A. M. MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill Book Co., New York, 1950.

[7] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, Princeton, 1943.

ESTIMATION OF THE MEAN AND STANDARD DEVIATION BY ORDER STATISTICS. PART II

By A. E. Sarhan

University of North Carolina

1. Introduction. In a previous paper [3], the best linear estimates of the mean and standard deviation for the rectangular, triangular, double exponential, and the exponential distributions were worked out. The best linear estimates were obtained by ranking the observations in ascending order and finding the best linear combination of them [2]. The variation of the coefficients in the estimates and the efficiencies of some other linear estimates were discussed.

This paper—which is a continuation of the previous one [3]—deals with three distributions: a U-shaped, a parabolic, and a skewed one. The same items were worked out for these distributions as for those in the previous paper. Also, a general idea of the natural sequence of the coefficients in the best linear estimate of the mean as the shape of the distribution undergoes change will be considered.

The mathematical formulae for this work will not be given as they are similar to those given in [3].

2. U-shaped population. The frequency distribution of a U-shaped population is

$$f(y) = \frac{3(y - \theta_1)^2}{2\theta_2^3}, \qquad \theta_1 - \theta_2 \leqq y \leqq \theta_1 + \theta_2 \tag{2.1}$$

where θ_1 is the mean and θ_2 is half the range. Standardizing the variable we get

$$f(x) = \tfrac{3}{2}x^2, \qquad -1 \leqq x \leqq +1. \tag{2.2}$$

The coefficients α_{1i} in the best linear estimates of the mean are given in Table I such that

$$\theta_1^* = \sum_{i=1}^{n} \alpha_{1i}\, y_{(i)}, \tag{2.3}$$

where $y_{(i)}$ is the ith ordered sample element.

Since

$$V(y) = \tfrac{3}{5}\theta_2^2, \tag{2.4}$$

we can estimate the standard deviation σ by $\sqrt{\tfrac{3}{5}}\ \theta_2^*$ and the coefficients can be adjusted to give the best linear estimate of the standard deviation σ^*. These adjusted coefficients for which

$$\sigma^* = \sum_{i=1}^{n} \alpha_{2i}\, y_{(i)} \tag{2.5}$$

are also shown in Table I.

Received August 23, 1954.

TABLE I

Coefficients in the best linear estimate of the mean and standard deviation, based on the order statistic $y_{(i)}$ in different populations of size n, for the mean $\theta_1^* = \sum_{i=1}^{n} \alpha_{1i} y_{(i)}$ and for the standard deviation $\sigma^* = \sum_{i=1}^{n} \alpha_{2i} y_{(i)}$

Sample size n and population	α_{11}	α_{12}	α_{13}	α_{14}	α_{15}	α_{21}	α_{22}	α_{23}	α_{24}	α_{25}
2										
U-shaped	.5000000	.5000000				−.9036961	+.9036961			
Parabolic	.5000000	.5000000				−.8695889	.8695889			
Skewed	.5000000	.5000000				−.8750000	.8750000			
3										
U-shaped	.5400000	−.0800000	.5400000			−.6024640	0	.6024640		
Parabolic	.4404762	.1190476	.4404762			−.5797259	0	.5797259		
Skewed	.4173734	.1539642	.4286624			−.5542721	−.0620262	.6162983		
4										
U-shaped	.5530728	−.0530728	−.0530728	.5530728		−.5218559	.0282930	−.0282930	.5218559	
Parabolic	.4078157	.0921843	.0921843	.4078157		−.4704178	−.0310128	.0310128	.4704178	
Skewed	.3703449	.1237432	.1145196	.3913123		−.4324793	−.0796294	−.0018219	.5139305	
5										
U-shaped	.5584792	−.0448589	−.0272406	−.0448589	.5584792	−.4890315	+.0318491	0	−.0318491	.4890315
Parabolic	.3862869	.0795401	.0683459	.0795401	.3862869	−.4105156	−.0392644	0	.0392644	.4105156
Skewed	.3387590	.1080812	.0903216	.0953199	.3675182	−.3655858	−.0791405	−.0305948	.0192066	.4561145

TABLE II

Variances of the best linear estimates of the mean and standard deviation in different populations ($\sigma = 1$)

Population and sample size	Variance of the estimate of mean	Variance of the estimate of standard deviation
2		
U-shaped	.5000000	.6333333
Rectangular	.5000000	.5000000
Parabolic	.5000000	.5123457
Triangular	.5000000	.5306132
Normal	.5000000	.57079
Double Exponential	.5000000	.7777778
3		
U-shaped	.2501299	.2161616
Rectangular	.3000000	.2000000
Parabolic	.3208101	.2220975
Triangular	.3293975	.2414966
Normal	.3333333	.27548
Double Exponential	.2947532	.4320999
4		
U-shaped	.1279837	.0955036
Rectangular	.2000000	.1111111
Parabolic	.2315500	.1335981
Triangular	.2443499	.1514217
Normal	.2500000	.18005
Double Exponential	.2077706	.2986242
5		
U-shaped	.0675462	.0470213
Rectangular	.1428371	.0714286
Parabolic	.1790064	.0925499
Triangular	.1934059	.1079590
Normal	.2000000	.13332
Double Exponential	.1584266	.2288250

The variances of the estimates of the mean and standard deviation are given in Table II. Furthermore, the relative efficiencies of the sample mean, median, and the midrange as estimates of the population mean are shown in Table III. Similarly the relative efficiencies of the range, the normal estimate, and Gini's estimate are also given in the same table. The efficiencies are calculated relative to the best linear estimate.

Table I shows that the two extreme values in the estimate of the mean have large weights while the middle elements have negative weights.

Comparing the efficiencies (Table III) of the estimates of the mean, we see that the midrange is more efficient than either the sample mean or the median. Again, the range as an estimate of standard deviation has a higher efficiency than either the normal or Gini's estimate. So, the midrange and the range (which are based on the two extreme values) can be used to estimate the popula-

TABLE III

Percentage efficiencies of certain estimates of the mean and standard deviation relative to BLE, *in different populations, from ordered samples of size n*

Population and sample size	Estimates of the mean			Estimates of standard deviation		
	$\bar{y}$	$\check{y}$	w	R	N	G
2						
U-shaped..............	100	100	100	100	100	100
Parabolic..............	100	100	100	100	100	100
Skewed..............	100	100	100	100	100	100
3						
U-shaped..............	75.04	30.57	98.77	100	100	100
Parabolic..............	96.24	60.25	98.83	100	100	100
Skewed..............	97.41	61.29	98.59	99.57	99.57	99.57
4						
U-shaped..............	51.19	23.95	97.16	99.61	86.34	90.99
Parabolic..............	92.62	65.27	97.35	99.81	52.80	97.73
Skewed..............	95.91	68.59	97.61	98.84	98.56	97.64
5						
U-shaped..............	33.77	9.36	95.39	99.01	77.50	69.27
Parabolic..............	89.50	49.56	95.91	99.55	96.03	91.49
Skewed..............	91.36	51.58	93.05	97.96	97.40	95.51

Here $\bar{y}$ denotes the sample mean, $\check{y}$ denotes median, w denotes the midrange, R denotes the range, N denotes the normal estimates, and G denotes the Gini's mean difference.

tion mean and standard deviation in this distribution for the sample sizes without great loss of accuracy.

3. Parabolic population. The frequency distribution of a parabolic population is

$$f(y) = \frac{6(y - \theta_1 + \frac{1}{2}\theta_2)(\theta_1 + \frac{1}{2}\theta_2 - y)}{\theta_2^3}, \qquad \theta_1 - \tfrac{1}{2}\theta_2 \leqq y \leqq \theta_1 + \tfrac{1}{2}\theta_2, \tag{3.1}$$

where θ_1 is the true mean and θ_2 is the range. Standardizing the variable we get

$$f(x) = 6x(1 - x), \qquad 0 \leqq x \leqq 1. \tag{3.2}$$

The coefficients α_{1i} in the best linear estimate of the mean (θ_1^*) are given in Table I.

Since

$$V(y) = \tfrac{1}{20}\theta_2^2, \tag{3.3}$$

we can estimate the standard deviation σ^* by $(1/\sqrt{20})\theta_2^*$ and the coefficients can be adjusted to give the best linear estimate of the standard deviation σ^*. These adjusted coefficients for which

$$\sigma^* = \sum_{i=1}^{n} \alpha_{2i}\, y_{(i)} \tag{3.4}$$

are given in Table I. The variances of the estimates of the mean and standard deviation are given in Table II.

Table III gives the percentage efficiencies of the different estimates relative to the best linear estimate. In the best linear estimate of the mean we find that the extreme values have higher weights while the middle elements have smaller positive weights (decreasing towards the middle).

For the given sample sizes, the midrange as an estimate of the population mean is shown to be more efficient than the sample mean (Table III), while the median has low efficiency. Furthermore, the range as an estimate of the standard deviation is more efficient than either the normal or the Gini's estimate as shown in Table III.

4. A skewed population. The frequency distribution of a skewed population is

$$\frac{12}{\theta_2}\left(\frac{y-\theta_1}{\theta_2}+\tfrac{2}{3}\right)^2\left(\tfrac{1}{3}-\frac{y-\theta_1}{\theta_2}\right), \qquad \theta_1-\frac{2\theta_2}{3}\leqq y\leqq \theta_1+\frac{2\theta_2}{3}, \tag{4.1}$$

where θ_1 is the true mode and θ_2 is the true range. Let

$$x=\frac{y-\theta_1}{\theta_2}+\tfrac{2}{3}, \tag{4.2}$$

to get

$$f(x) = 12x^2(1-x). \tag{4.3}$$

Since the population mean is $\theta_1 - \theta_2/15$, and the population standard deviation is $\theta_2/5$, then the coefficients can be adjusted to give the estimates for the mean μ and the standard deviation σ. These can be obtained from

$$\mu^* = \theta_1^* - \theta_2^*/15, \tag{4.4}$$

$$\sigma^* = \theta_2^*/5. \tag{4.5}$$

The adjusted coefficients in the BLE of the mean μ^* and standard deviation σ^* are given in Table I. The efficiencies of the estimates are given in Table III.

In this case, again, we find that the two extreme sample elements have the greatest numerical weights in the BLE while the other values have smaller weights. It is of interest to see that the least sample value (the extreme value on the side of the long tail) has a smaller coefficient than the largest sample value (the other extreme on the side of the shorter tail). This is to be expected since extreme values from the longer tail occur more often and tend to upset the estimate. It throws some light on the effect of the shape of the distribution or the length of its tails on the coefficients of the BLE. This is not the only relation, however, and the nature of the general relation is not yet well known.

The midrange has a higher efficiency for the given sample sizes than that of the sample mean while the median has a lower efficiency. Again, the range has a higher efficiency than either the normal or the Gini's estimate.

5. Coefficients in the BLE of the mean for symmetric distributions. We have seen in [3], and in the previous sections that the coefficients in the best linear estimate of the mean vary as the parent distribution undergoes change. It is of interest to notice the sequence of this variation. The sample elements may have equal weights or smaller weights at the middle than at the tail, or zero weights at the middle and equal weights at the extremes or large weights on the tails and negative weights in the middle. There is a sequence in which the middle elements are to be equally weighted, zero weighted, and negatively weighted.[1] It seems that the full sequence is missing its natural extension and the complete sequence should read:

(a) negative weights in the middle and large positive weights at the tails,
(b) zero weights in the middle and equal weights at the tails,
(c) less weights in the middle than at the tails,
(d) equal weights throughout,
(e) more weight in the center and less weights in tails, but all positive weights,
(f) middle observations receive all the weight, others nothing,
(g) middle observations receive more than unity and tails take on negative weights.

This is the sequence which might be anticipated. The results show that (a) is U-shaped; (b) is rectangular; (c) is triangular or parabolic; (d) is normal; (e) is double exponential; (f) is the case where the median gets all the weight, which is like a double exponential but not exactly. For (g) the author does not know any example at this time, i.e., a distribution where it would be best to estimate the mean by giving the middle element a weight greater than one and to give the elements on the tails some negative weights. This represents, however, a natural continuity in the sequence.

6. The variances of the best linear estimates. Table II gives the variances of the best linear estimates of the mean and standard deviation in different symmetric distributions with $\sigma = 1$. The variances of the estimates of the normal population are obtained from Tables 5 and 6 in [1] calculated to five decimal places.

The table shows that the variance of the best linear estimate of the mean of a U-shaped population (for $n > 2$) is the least among the given distributions. This raises the theoretical problem of finding the distribution whose mean can be estimated with the least variance.

The same table shows also that the variance of the estimate of the mean increases gradually from the case of the U-shaped distribution to the rectangular, to the parabolic, to the triangular, and then to the normal. The variance of the estimate of the mean then decreases in the case of double exponential.

As to the variance of the estimate of standard deviation, the same table shows that the variance of the estimate increases from the rectangular to the

[1] The author wishes to thank Professor Frederick Mosteller for directing his attention to this particular sequence.

parabolic, to the triangular, to the normal, and then to the double exponential. For the U-shaped distribution, the variance of the estimate is greater than that of the rectangular for $n = 2$ and 3. For $n = 4$ and 5, the variance becomes smaller than that of the rectangular. However, working out the estimates and their variances for $n = 6$ and 7, it has been found that the variance of the estimate of standard deviation for the U-shaped becomes progressively smaller than that of the rectangular. So it seems to the author that as n increases, the variance of the estimate of the standard deviation of the U-shaped distribution tends to be the least among the given distributions.

The author wishes to acknowledge the kind help of Dr. B. G. Greenberg, under whose direction this work was done.

REFERENCES

[1] A. K. Gupta, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 260–273.

[2] E. H. Lloyd, "Least squares estimation of location and scale parameters using order statistics," *Biometrika*, Vol. 39 (1952), pp. 88–95.

[3] A. E. Sarhan, "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 317–328.

PROBABILITY OF INDECOMPOSABILITY OF A RANDOM MAPPING FUNCTION[1]

By Leo Katz[2]

Michigan State University

Summary. Consider a finite set Ω of N points and a single-valued function $f(x)$ on Ω into Ω. In case the mapping is one-to-one, it is a permutation of the points of Ω; we shall be concerned with more general mappings. Any mapping function effects a decomposition of the set into disjoint, *minimal*, non-null invariant subsets, as $\Omega = \omega_1 + \omega_2 + \cdots + \omega_k$, where $f(\omega_i) \subset \omega_i$ and $f^{-1}(\omega_i) \subset \omega_i$. These subsets have been referred to as trees and as components of the mapping; we shall say that f, as above, decomposes the set into k components.

Metropolis and Ulam [1] defined a *random* mapping by a uniform probability distribution over the Ω^{Ω} sample points of $f(x)$ and posed the problem of finding the expected number of components. Kruskal [2] subsequently solved this problem. In this paper, we consider a related problem, namely, what is the probability that a random mapping is indecomposable, i.e., that the minimal non-null set ω for which $f(\omega) = \omega$ and $f^{-1}(\omega) = \omega$, is the whole set $\omega = \Omega$?

This problem is solved in general, as is, also, an analogous problem for a specialized random mapping of some interest in social psychology. Finally, we examine the asymptotic behavior of these probabilities.

1. Indecomposability of a random mapping. A single-valued mapping specifies, for each point P_i, its image point P_{j_i}, $j_i = 1, 2, \cdots, N$ (a point may map into itself). A *random* mapping assigns, independently, to each P_i one of the image points P_j, $j = 1, 2, \cdots, N$, with equal probability $1/N$. The sample space consists of the N^N possible mappings, with uniform probability distribution. To each mapping is associated a value of the random variable k, $k = 1, 2, \cdots, N$, the number of components. Those for which $k = 1$ are indecomposable. We shall require, first, a characterization of the property of indecomposability, second, a disjunctive and exhaustive categorization of those mappings which possess this property, and, finally, an enumeration scheme within each category.

In order to obtain a suitable characterization of indecomposability, we consider that a single-valued mapping function takes any point of the (finite) set into a second, the second into a third, etc., until, at some stage, a point is taken into an earlier member of the sequence. At this stage, a cycle is formed; the length of the cycle is the number of repetitions of the mapping required to

Received November 8, 1954.

[1] Presented at the meeting of the Institute of Mathematical Statistics at Berkeley, California, December 27–30, 1954.

[2] Work done under the sponsorship of the Office of Naval Research.

take any point of the cycle into itself. No point of a cycle can be mapped on any point not of the cycle, but a point not of a cycle may map through a chained sequence into a point of the cycle. Thus, a *component* of a mapping consists of precisely one cycle, together with cycle-free chains terminating at points of the cycle. This provides the required characterization:

CHARACTERIZATION. A mapping function is indecomposable if and only if it generates only one cycle.

Next, we may categorize indecomposable mappings according to the length m of the cycle contained in them. Finally, we subcategorize m-cycle indecomposable mappings into sets according as the noncyclic elements are arranged with n_j requiring j stages to be mapped into the cycle, $j = 1, 2, 3, \cdots$. This subclassification corresponds to the nonzero, p-part, partitions of $(N - m)$, with p arbitrary.

We now view the indecomposable mapping as a directed graph, more precisely, as a tree rooted in an m-cycle. The directed joins, one emanating from each point, represent the mapping from point to image. In the following section, we shall consider that the graph of an indecomposable mapping consists of a central m-cycle, a first orbit of n_1 points connected by one-chains to the cycle, a second orbit of n_2 points connected by one-chains to the points of the first orbit and, hence, by two-chains to the points of the cycle, etc. An example of such an indecomposable mapping with $m = 6$, $n_1 = 4$, $n_2 = 3$ is given in Figure 1, below.

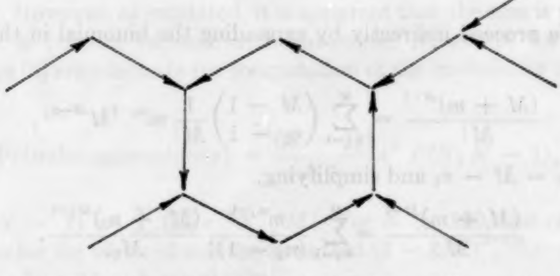

FIG. 1. Example of Mapping $m = 6$, $n_1 = 4$, $n_2 = 3$

2. Probability that mapping is indecomposable. We proceed formally, at first, to express the probability of indecomposability as the sum of compound probabilities that the mapping produces exactly one cycle *and* the cycle is of length m. These, in turn, are expressed as the sums of probabilities that the remaining $M = N - m$ are arranged in nonempty orbits of $n_1, n_2, \cdots, n_p$, respectively, for all possible such arrangements. It is convenient to give special treatment to the number m in the cycle itself. Consider the event $E_m(n_1, n_2, \cdots, n_p)$ that a random mapping is indecomposable with parameters $m, n_1, \cdots, n_p$. The

probability of this event is the curiously linked expression

$$P\{E_m(n_1, n_2, \cdots, n_p)\}$$

(1)
$$= \binom{N}{m, n_1, \cdots, n_p} \frac{(m-1)!}{N^m} \left(\frac{m}{N}\right)^{n_1} \left(\frac{n_1}{N}\right)^{n_2} \cdots \left(\frac{n_{p-1}}{N}\right)^{n_p},$$

where $\binom{a}{b,c,d,\ldots}$ is the multinomial coefficient. The factorial in the second factor of the right member represents the number of distinct cyclical arrangements possible among the m points of the inner cycle; succeeding factors represent the possibilities of joins of points in an orbit to points of the next interior orbit.

With slight rearrangement of (1), the probability we seek is given by

(2)
$$\Pr\{\text{indecomposability}\} = \sum_{m=1}^{N} \frac{N!}{N^N} \left\{ \sum_p \sum_{[M]_p} \frac{1}{m} \frac{m^{n_1} n_1^{n_2} \cdots n_{p-1}^{n_p}}{n_1! n_2! \cdots n_p!} \right\},$$

where $[M]_p$ stands for the collection of nonempty, p-part partitions

$$(n_1, n_2, \cdots, n_p)$$

of M.

We now evaluate the expression in braces in (2) by the following lemma.

LEMMA.

$$\sum_p \sum_{[M]_p} \frac{1}{m} \frac{m^{n_1} n_1^{n_2} \cdots n_{p-1}^{n_p}}{n_1! n_2! \cdots n_p!} = \frac{N^{M-1}}{M!}$$

where $N = M + m$ and $[M]$ is the class of nonzero distinct partitions $(n_1, \cdots, n_p)$ of M.

PROOF.[3] We proceed indirectly by expanding the binomial in the right member as

(3)
$$\frac{(M+m)^{M-1}}{M!} = \sum_{n_1=1}^{M} \binom{M-1}{n_1-1} \frac{1}{M!} m^{n_1-1} M^{M-n_1},$$

or, letting $M_1 = M - n_1$ and simplifying,

(3a)
$$\frac{(M+m)^{M-1}}{M!} = \sum_{n_1=1}^{M} \frac{m^{n_1-1}}{(n_1-1)!} \frac{(M_1+n_1)^{M_1-1}}{M_1!}.$$

We note that the second factor in the summand of (3a) is of the same type as the left member and may be similarly expanded. Letting $M_i = M_{i-1} - n_i$, $i = 2, 3, \cdots$, we obtain, by iteration of (3a),

(4)
$$\frac{(M+m)^{M-1}}{M!} = \sum_{n_1=1}^{M} \frac{m^{n_1-1}}{(n_1-1)!} \sum_{n_2=1}^{M_1} \frac{n_1^{n_2-1}}{(n_2-1)!} \cdots \sum_{n_p=1}^{M_{p-1}} \frac{n_{p-1}^{n_p-1}}{(n_p-1)!} n_p^{-1},$$

with p arbitrary. But the summations in the right member are equivalent to the sum over all p and nonzero p-part partitions of M and the summand is that of the lemma, thus proving the lemma.

[3] This short proof is partly due to J. S. Frame.

The lemma and (2), upon changing the index of summation to $M = N - m$, gives immediately the principal theorem:

THEOREM. *The probability that a random mapping on N points is indecomposable is*

$$[(N-1)!/N^N] \sum_{M=0}^{N-1} N^M/M!.$$

3. Hollow mapping. One realization of near-random mapping occurs in sociometric testing. When, for example, N individuals in a group are each asked to indicate which one of the others is his best source of information, the result is such a mapping except that, if no individual is permitted to name himself, the mapping is "hollow" in the sense that the matrix representation of the graph has diagonal elements identically vanishing. If, otherwise, selection is random, the probability of equation (1) is modified for this case by replacing each N in the denominator by (N-1) and taking the outer summation of equation (2) from $m = 2$ to $m = N$. With these adjustments, we have the following corollary.

COROLLARY. *The probability that a hollow random mapping on N points is indecomposable is*

$$[(N-1)!/(N-1)^N] \sum_{M=0}^{N-2} N^M/M!.$$

4. Computation and asymptotic probability. The probabilities of indecomposability, of the theorem and the corollary above, might be expressed in more compact form. However, as exhibited, it is apparent that the sum is a cumulative probability of a Poisson variable with parameter N, except for a constant. Molina's tables [3] are adequate for computation of the probability through $N = 100$. Thus,

$$\text{(5)} \qquad \Pr\{\text{indecomposability}\} = \frac{(N-1)!}{N^N} e^N P(N; N-1),$$

where $P(N; N-1) = \sum_{M=0}^{N-1} e^{-N} N^M/M!$. For $N > 100$, use of the Stirling approximation for the factorial and the facts that $(1 - 1/N)^{N-1/2} = e^{-1} + O(N^{-2})$ and that $P(N; N-1) \to \frac{1}{2}$, we obtain

$$\text{(6)} \qquad \Pr\{\text{indecomposability}\} = \left(\frac{\pi}{2N}\right)^{1/2}, \quad N \text{ large.}$$

Similarly, using the corollary, we have

$$\text{(5h)} \qquad \Pr\{\text{indecomposability} \mid \text{hollow}\} = \frac{(N-1)!}{(N-1)^N} e^N P(N; N-2),$$

$$\text{(6h)} \qquad \Pr\{\text{indecomposability} \mid \text{hollow}\} = e\left(\frac{\pi}{2(N-1)}\right)^{1/2}, \quad N \text{ large.}$$

TABLE I

Probabilities of Indecomposability of a random mapping function in the general and hollow cases

N	P{I|G}	P{I|H}	N	P{I|G}	P{I|H}
2	.75000	1.00000	26	.23372	.54135
3	.62963	1.00000	28	.22562	.52574
4	.55469	.96296	30	.21831	.51148
5	.50208	.92188	32	.21169	.49837
6	.46245	.88320	34	.20564	.48628
7	.43116	.84816	36	.20009	.47507
8	.40563	.81671	38	.19497	.46463
9	.38426	.78844	40	.19023	.45488
10	.36602	.76294	45	.17976	.43308
11	.35022	.73983	50	.17086	.41426
12	.33634	.71878	55	.16318	.39780
13	.32403	.69950	60	.15646	.38322
14	.31300	.68176	65	.15052	.37020
15	.30305	.66539	70	.14521	.35847
16	.29402	.65019	75	.14043	.34782
17	.28576	.63605	80	.13610	.33810
18	.27817	.62284	85	.13215	.32918
19	.27117	.61047	90	.12853	.32096
20	.26468	.59885	95	.12519	.31334
22	.25301	.57757	100	.12210	.30626
24	.24333	.55853	Large N	$(\pi/2N)^{1/2}$	$e(\pi/2(N-1))^{1/2}$

The most interesting feature of this last result is that the probability in the hollow case remains substantially larger than in the general case as N increases. This runs counter to standard sociometric folklore, which holds that the hollow model may be uniformly replaced by the general model with small error when N is large.

Both probabilities approach zero fairly slowly (as $N^{-1/2}$). Table I presents the exact probabilities as computed from (5) and (5a).[4]

5. Notes on related work. After the present paper had been prepared, David Blackwell called the attention of the author to an unpublished memorandum by Rubin and Sitgreaves [4]. In the memorandum, different methods are used to obtain the theorem of Section 2 of this paper; the hollow mapping case is not considered.

Using methods of this paper, Jay E. Folkert and the author have obtained and will publish the probability distributions of the numbers of components of single-valued and of multiple-valued mappings in the subcases in which mapping is arbitrary or hollow. The distribution for the single-valued, arbitrary case is given also in the memorandum cited above.

---

[4] The author is indebted to Mr. William L. Harkness for these computations.

REFERENCES

[1] N. Metropolis and S. Ulam, "A property of randomness of an arithmetical function," *Amer. Math. Monthly*, Vol. 60 (1953), pp. 252–253.

[2] M. D. Kruskal, "The expected number of components under a random mapping function," *Amer. Math. Monthly*, Vol. 61 (1954), pp. 392–397.

[3] E. C. Molina, *Poisson's Exponential Binomial Limit*, D. Van Nostrand, 1942.

[4] H. Rubin and R. Sitgreaves, "Probability Distributions Related to Random Transformations on a Finite Set," Tech. Rep. #19A, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.

NOTES

A NECESSARY AND SUFFICIENT CONDITION FOR ADMISSIBILITY

By Charles Stein

Stanford University

1. Summary. In Section 2 we give the usual definition for admissibility of a strategy in a two-person zero-sum game, and obtain a simple sufficient condition for admissibility of a strategy for the second player which is hardly more than a formal statement of a procedure frequently used in proving admissibility. In Section 3 we introduce the notion of strict admissibility, which is slightly stronger than admissibility, but equivalent to it in the case where the space of strategies of the second player is weakly compact in the sense of Wald. We then obtain a necessary and sufficient condition for strict admissibility, in the form of a condition on the upper values of a sequence of games associated with the original game. In Section 4 we show that, under the additional condition that the minimax theorem holds for certain associated games, the condition of Section 2 is necessary as well as sufficient. The results have a formal resemblance to those of Hodges and Lehmann [4].

2. Introduction. Let A and B be sets and K a real-valued function on $A \times B$ such that for every $a \,\varepsilon\, A$

$$\rho(a) = \inf_{b \varepsilon B} K(a,b) > -\infty. \tag{1}$$

Following von Neumann [1], we refer to the triple (A, B, K) as a two-person zero-sum game, having in mind the situation where the first player chooses an element a of A, the second player an element b of B, the two choices being made simultaneously, and then the second player pays the first player the amount $K(a, b)$. The set B is partially ordered by the relation $\leqq$ where $b_1 \leqq b_2$ means that, for every $a \,\varepsilon\, A$,

$$K(a, b_1) \leqq K(a, b_2). \tag{2}$$

If this holds we say that b_1 *is better than* b_2. If, for all a,

$$K(a, b_1) = K(a, b_2), \tag{3}$$

we write $b_1 \approx b_2$ and say that b_1 *is equivalent to* b_2. If $b_1 \leqq b_2$ but not $b_1 \approx b_2$, we write $b_1 < b_2$ and say that b_1 *is strictly better than* b_2. We say that b_1 is *admissible* if there exists no b_2 strictly better than b_1.

We shall need a few more definitions before we can indicate the principal result of this paper. The strategy b_1 is said to be *ϵ-Bayes* with respect to $a \,\varepsilon\, A$ if

$$K(a, b_1) \leqq \inf_b K(a, b) + \epsilon, \tag{4}$$

Received November 17, 1954.

and to be *Bayes* if this holds for $\epsilon = 0$. If $\mathcal{A}$ is a σ-algebra of subsets of A such that for each b, $K(\cdot, b)$ is $\mathcal{A}$ measurable, and Ξ is a convex set of probability measures on $\mathcal{A}$, including at least all those measures, denoted by $[a]$, concentrated at a single point $a \varepsilon A$, then the game (Ξ, B, K') with

$$K'(\xi, b) = \int K(a, b)\, d\xi(a) \tag{5}$$

will be called a *convex extension* of K. In order to make sure that this integral is defined we must make an additional assumption on K, and we shall assume K bounded below. A reasonable alternative might be the condition symmetric to (1), that is,

$$\sup_a K(a, b) < \infty \qquad \text{for all } b. \tag{6}$$

THEOREM 1. *If b_1 is such that for every $a_1 \varepsilon A$ and $\epsilon > 0$ there exists $\xi \varepsilon \Xi$ and $\delta > 0$ such that b_1 is $\epsilon\delta$-Bayes with respect to $(1 - \delta)\xi + \delta[a_1]$, then b_1 is admissible.*

PROOF. Suppose b_1 is not admissible. Then there must exist b_2 which is strictly better than b_1, that is,

$$K(a, b_2) \leqq K(a, b_1) \tag{7}$$

for all a with strict inequality for some a, say a_1. By assumption, there exists $\delta > 0$ and $\xi \varepsilon \Xi$ such that

$$\begin{aligned} K'((1 - \delta)\xi + \delta[a_1], b_1) &\leqq \inf_b K'((1 - \delta)\xi + \delta[a_1], b) + \epsilon\delta \\ &\leqq K'((1 - \delta)\xi + \delta[a_1], b_2) + \epsilon\delta \\ &\leqq (1 - \delta)K(\xi, b_1) + \delta K(a_1, b_2) + \epsilon\delta \end{aligned}$$

so that $K(a_1, b_1) \leqq K(a_1, b_2) + \epsilon$. Since ϵ is arbitrary, $K(a_1, b_1) \leqq K(a_1, b_2)$, which contradicts the hypothesis that (7) holds with strict inequality at a_1.

This theorem essentially follows the reasoning used by Blyth [2] and other authors in proving admissibility. In Section 4, assuming weak compactness of B in the sense of Wald [3], and assuming the minimax theorem to hold for a class of games associated with K', we shall show that this condition is also necessary. The set B is said to be *weakly compact* with respect to K in the sense of Wald if, for every sequence $\{b_i\}$, there exists b_0 and a subsequence $\{b_{i_j}\}$ such that

$$\varliminf_{j \to \infty} K(a, b_{i_j}) \geqq K(a, b_0). \tag{8}$$

We observe that, by Fatou's Lemma, if B is weakly compact with respect to K, and K is bounded below, then B is weakly compact with respect to K'.

The necessity of the condition of Theorem 1 could perhaps be proved more quickly without the intermediate results of Section 3. However, the necessary and sufficient condition, valid under much weaker conditions which we obtain there, is likely to be of some interest.

3. A necessary and sufficient condition for admissibility. In this section we shall use the notation and assumptions of Section 2 through (2.4), and also the definition of weak compactness (2.8). We shall also need the notion of strict admissibility, slightly stronger than admissibility. The strategy b_1 is said to be *strictly admissible* if for every $a_1 \varepsilon A$ and $\epsilon > 0$, there exists $\delta > 0$ such that, for every b for which $K(a_1, b) \leqq K(a_1, b_1) - \epsilon$, there exists a such that $K(a, b) \geqq K(a, b_1) + \delta$. It is clear that strict admissibility implies admissibility.

THEOREM 2. *If b_0 is admissible and B is weakly compact with respect to K, then b_0 is strictly admissible.*

PROOF. Suppose B is weakly compact in the sense of Wald and b_0 is not strictly admissible. Then for some $a_0 \varepsilon A$ and some $\epsilon > 0$ there exists a sequence $\{b_i\}$ such that

$$K(a_0, b_i) \leqq K(a_0, b_0) - \epsilon \quad \text{for all } i = 1, 2, \cdots, \tag{1}$$

$$\limsup_{i\to\infty} \sup_{a \varepsilon A} [K(a, b_i) - K(a, b_0)] \leqq 0. \tag{2}$$

By the assumption of weak compactness there exists a subsequence $\{b_{i_j}\}$ and an element b' such that

$$\varliminf_{j\to\infty} K(a, b_{i_j}) \geqq K(a, b') \quad \text{for all} \quad a. \tag{3}$$

It follows that

$$\begin{aligned} \sup_{a \varepsilon A} [K(a, b') - K(a, b_0)] &\leqq \sup_{a \varepsilon A} [\varliminf_{j\to\infty} K(a, b_{i_j}) - K(a, b_0)] \\ &= \sup_{a \varepsilon A} \varliminf_{j\to\infty} [K(a, b_{i_j}) - K(a, b_0)] \\ &\leqq \varliminf_{j\to\infty} \sup_{a \varepsilon A} [K(a, b_{i_j}) - K(a, b_0)] \leqq 0. \end{aligned} \tag{4}$$

Similarly,

$$K(a_0, b') \leqq K(a_0, b_0) - \epsilon, \tag{5}$$

so that b' is strictly better than b_0. Thus b_0 is not admissible.

THEOREM 3. *In order that b_0 be strictly admissible, it is necessary and sufficient that for every a_0*

$$\varlimsup_{\gamma\to\infty} \inf_b \sup_a \{K(a_0, b) - K(a_0, b_0) + \gamma[K(a, b) - K(a, b_0)]\} \geqq 0. \tag{6}$$

In order to simplify the writing we assume without essential loss of generality that, for all a, $K(a, b_0) = 0$. Then (6) becomes

$$\varlimsup_{\gamma\to\infty} \inf_b \sup_a \{K(a_0, b) + \gamma K(a, b)\} \geqq 0. \tag{7}$$

Also a_0 may be taken as fixed throughout the proof.

PROOF OF NECESSITY. Suppose b_0 strictly admissible and let δ_ϵ be the δ whose

existence is asserted in the definition of strict admissibility. Let S_ϵ be the set of all b such that

$$K(a_0, b) < -\epsilon \tag{8}$$

and S'_ϵ its complement. Then

$$\begin{aligned}
&\varlimsup_{\gamma\to\infty} \inf_b \sup_a [K(a_0, b) + \gamma K(a, b)] \\
&\qquad = \varliminf_{\gamma\to\infty} \min \{ \inf_{b\epsilon S_\epsilon} [K(a_0, b) + \gamma \sup_a K(a, b)], \\
&\inf_{b\epsilon S'_\epsilon} [K(a_0, b) + \gamma \sup_a K(a, b)]\} \qquad (9)\\
&\qquad \geqq \lim_{\gamma\to\infty} \min \{\rho(a_0) + \gamma\delta_\epsilon, \quad -\epsilon + \gamma \inf_b \sup_a K(a, b)\} \geqq -\epsilon.
\end{aligned}$$

The last step follows from the fact that, by the admissibility of b_0, for every b there exists a such that $K(a, b) \geqq 0$, so that

$$\inf_b \sup_a K(a, b) \geqq 0.$$

Since ϵ was arbitrary, this completes the proof of necessity.

PROOF OF SUFFICIENCY. Supposing that (7) holds we have for every $\epsilon > 0$,

$$\begin{aligned}
0 &\leqq \varlimsup_{\gamma\to\infty} \inf_b \sup_a [K(a_0, b) + \gamma K(a, b)] \\
&\leqq \varlimsup_{\gamma\to\infty} \inf_{b\epsilon S_\epsilon} \sup_a [K(a_0, b) + \gamma K(a_0, b)] \\
&\leqq \varlimsup_{\gamma\to\infty} [\sup_{b\epsilon S_\epsilon} K(a_0, b) + \gamma \inf_{b\epsilon S_\epsilon} \sup_a K(a, b)] \qquad (10)\\
&= \lim_{\gamma\to\infty} [-\epsilon + \gamma \inf_{b\epsilon S_\epsilon} \sup_a K(a, b)].
\end{aligned}$$

Consequently, there exists $\gamma_\epsilon > 0$ such that

$$-\tfrac{1}{2}\epsilon \leqq -\epsilon + \gamma_\epsilon \inf_{b\epsilon S_\epsilon} \sup_a K(a, b). \tag{11}$$

Thus the definition of strict admissibility is satisfied with $\delta = \frac{1}{2}\epsilon/\gamma_\epsilon$.

The proof shows that (6) could have been stated with $\overline{\lim}$ replaced by $\underline{\lim}$ or by lim, or with $\geqq$ replaced by $=$, or both.

COROLLARY. *If (6) holds for all a_0, then b_0 is admissible. If B is weakly compact, then the converse holds.*

This is an immediate consequence of Theorems 2 and 3.

4. Admissibility in the presence of the minimax theorem. In this section, we suppose K is bounded below and possesses a convex extension (Ξ, B, K') as described around (2.5). We shall also suppose the minimax theorem applies in (2.6) when K is replaced by K', that is,

$$\begin{aligned}
&\inf_b \sup_\xi \{K(a_0, b) - K(a_0, b_0) + \gamma[K'(\xi, b) - K'(\xi, b_0)]\} \qquad (1)\\
&\qquad = \sup_\xi \inf_b \{K(a_0, b) - K(a_0, b_0) + \gamma[K'(\xi, b) - K'(\xi, b_0)]\}.
\end{aligned}$$

THEOREM 4. *Under the above conditions, in order that b_0 be admissible it is necessary and sufficient that for every a_0 and every $\epsilon > 0$ there exist $\xi_1 \varepsilon \Xi$ and $\delta > 0$ such that b_0 is $\epsilon\delta$-Bayes with respect to* $(1 - \delta)\xi_1 + \delta[a_0]$.

PROOF. Using the fact that

$$\sup_{\xi} [K'(\xi, b) - K'(\xi, b_0)] = \sup_{a} [K(a, b) - K(a, b_0)] \tag{2}$$

together with (1), we find that (3.6) is equivalent to

$$\lim_{\gamma\to\infty} \sup_{\xi} \inf_{b} \{K(a_0, b) - K(a_0, b_0) + \gamma[K'(\xi, b) - K'(\xi, b_0)]\} \geqq 0. \tag{3}$$

If we let $\delta = 1/(\gamma + 1)$ and use the fact that

$$\frac{\gamma}{1+\gamma} K'(\xi, b) + \frac{1}{1+\gamma} K'([a_0], b) = K'\left(\frac{\gamma}{1+\gamma}\xi + \frac{1}{1+\gamma}[a_0], b\right), \tag{4}$$

we find that (3) is equivalent to

$$\lim_{\delta\downarrow 0} \frac{1}{\delta} \sup_{\xi} \inf_{b} [K'((1 - \delta)\xi + \delta[a_0], b) - K'((1 - \delta)\xi + \delta[a_0], b_0)] \geqq 0. \tag{5}$$

This is equivalent to the assertion that for every $\epsilon > 0$ there exist $\delta > 0$ and $\xi_1 \varepsilon \Xi$ such that

$$\inf_b K'((1 - \delta)\xi_1 + \delta[a_0], b) \geqq K'((1 - \delta)\xi_1 + \delta[a_0], b_0) - \epsilon\delta. \tag{6}$$

The theorem follows immediately.

REFERENCES

[1] J. VON NEUMANN, "Zur Theorie der Gesellschaftsspiele," *Math. Ann.*, Vol. 100 (1928), pp. 295–320.
[2] C. BLYTH, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 22–42.
[3] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.
[4] J. HODGES, AND E. LEHMANN, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 396–407.

Note added in proof. I believe this theorem to be potentially useful, but cannot now give any non-trivial examples. Attempts to apply the sufficiency often run into analytic difficulties. The necessity was useful heuristically in the recognition of the inadmissibility of the usual estimate of the mean of a multivariate normal distribution of dimension greater than or equal to 3. (Abstract in *Ann. Math. Stat.*, Vol. 26 (1955), p. 157; to appear in the *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*). A result similar to Theorem 4 has been obtained independently by LeCam.

A BIVARIATE SIGN TEST

By J. L. Hodges, Jr.

University of California, Berkeley

1. Introduction. The sign test has proved to be a very useful means for judging the significance of treatments. Suppose that on each of n individuals (or pairs of individuals) measurements are made under two conditions, for example, before and after treatment (or on a treated and a control subject). Denote the two measurements for the ith individual (or pair of individuals) by x_i and x_i'. We formulate the null hypothesis that x_i and x_i' are identically and independently distributed, but wish to make no assumption concerning relations between the distributions of $x_1, x_2, \cdots, x_n$, nor concerning relations between those of $x_1', x_2', \cdots, x_n'$, save that each set is independent. The alternative to the null hypothesis is that the second measurements x_i' are generally shifted, with respect to the first measurements x_i, in the same direction for all (or most) of the individuals. The test is carried out by counting the number S of the differences $x_i' - x_i$ which have positive signs. Under the null hypothesis, S is binomially distributed with $p = \frac{1}{2}$, assuming there are no cases with $x_i' = x_i$, or that such cases of equality are broken randomly. Under the alternative, S would tend to have large values if the second measurements are generally increased relative to the first, small values if they are decreased. We may then reject for large S, small S, or either, according to the alternative against which we wish the test to have power. The great advantage of the test, aside from its simplicity, is the generality of the conditions under which it is valid.

The present paper proposes a bivariate analog of the two-sided sign test, which can be applied when two quantities are measured on each individual. We now have measurements x_i and y_i in a first circumstance, x_i' and y_i' in a second. Do the $4n$ measurements justify our concluding that the two circumstances differ? The null hypotheses is that the bivariate distribution for (x_i, y_i) is identical with that for (x_i', y_i'), and that these vectors are independent. The alternative of interest is that in the second circumstance the bivariate distribution has been shifted relative to the first, in generally the same direction for all individuals. The direction of this possible shift is, however, unknown.

To illustrate, suppose we measure blood pressure and blood sugar before and after treatment with a new drug on a number of individuals. We wish to know whether the drug influences these quantities, but have no preconceived notion concerning the direction or relative amount of the influence on either quantity, should it exist. The joint distribution of the quantities has an unknown form, and is presumably different in different individuals. The quantities are presumably dependent, but in an unknown way.

If we knew the direction of a possible shift, it would be easy to reduce our problem to the sign test. We could simply project the vectors of differences

Received October 14, 1954.

$(x_i' - x_i, \ y_i' - y_i)$ onto the given direction, and count the number S of projected vectors having the given sense. Our problem arises just because we do not have a given direction, but must derive one from the data.

The idea of the proposed test is to consider all possible directions, and calculate S for each. Let M be the maximum of the values thus calculated. We shall use M as our test statistic, rejecting the null hypothesis if M is too large. That is, we shall judge that a shift has occurred if there is *some* direction in which most of the measurement pairs have shifted; we shall judge that no shift has occurred if the shifts are in various directions with no heavy concentration.

The distribution theory for M under the null hypothesis is worked out in the following sections. Presumably it would be desirable to generalize the proposed test to more than two quantities. The multivariate analog of the statistic M is easily seen, though in more than three dimensions it would be difficult to compute M from the sample, and its null distribution might be troublesome.

2. Reduction to a combinatorial problem. We shall suppose that none of the n vectors $(x_i' - x_i, \ y_i' - y_i)$ lies on the same line, and take the n lines on which these vectors lie as given, with all probability calculations conditional on the given lines. Under the null hypothesis, the distribution of $(x_i' - x_i, \ y_i' - y_i)$ is the same as that of $(x_i - x_i', \ y_i - y_i')$, so that there is probability $\frac{1}{2}$ for the ith vector to be oriented in each of its two possible senses. As the n vectors are independent, we conclude that the 2^n possible orientations of the vectors are all equally likely.

It is easily seen that the value of M for a given set of orientations is independent of the angles between the lines and of the lengths of the vectors. Therefore, for simplicity we may suppose that the lines are equally spaced and the vectors all are of unit length. We imagine a circle on whose circumference $2n$ equally spaced loci are given. We are to distribute n plus signs and n minus signs among these loci, subject to the condition that diametrically opposed signs are opposite in sense. We shall call such an arrangement a *cycle*. We think of a cycle as being rotatable about its center into $2n$ positions, each being itself a cycle. For each position we count the number s of positive signs among the n uppermost signs; m is the maximum of the $2n$ values of s thus obtained. Our problem is to count the cycles having a given value of m.

It is clear that $\frac{1}{2}n \leqq m \leqq n$. We shall denote $n - m$ by k; thus k is the smallest number of minus signs which can be uppermost. The operation of rotation carries one cycle into another, generating equivalence classes of cycles. The largest possible class has $2n$ members. Smaller classes are possible, since there may exist cycles which are carried into themselves by a rotation through r positions, $0 < r < 2n$. However, the smallest such r must be of the form $rc = 2n$ where $3 \leqq c$ is odd (since opposite signs are of opposite sense); thus cycles in an equivalence class smaller than $2n$ will have $k \geqq n/3$. As our interest is primarily in the tail of the distribution (k small), we shall simplify by restricting $k < n/3$, whence we can assume every class to have $2n$ members.

To count the classes, we shall select from each class a representative member, called the *pattern* for the class. This member is the unique one which satisfies two conditions, which can be expressed in terms of the n uppermost signs. These signs are arranged in a semi-circle, and we are particularly interested in the signs forming a consecutive set of fewer than n signs at either extreme of the semi-circle; we call such a set a (right or left) *tail*. The two conditions are:

(a) There is no right tail in which there is a majority of minus signs.

(b) There is no left tail in which the plus signs are not in the majority.

The conditions serve to insure that the pattern has the maximum number m of positive signs uppermost; if it were possible to rotate it into a position with more positive signs uppermost, there would have to be a tail with a majority of minus signs. The conditions also insure that only one pattern is selected from each class; if there were two representatives of the class, (i.e., a cycle appearing in two positions) one of these would contradict condition (a). In general it is not true that every class has a member satisfying these conditions (consider a cycle with alternating signs), but it is true under the restriction $k < n/3$.

3. Counting the patterns. We may obtain a formula for the number $P(n, k)$ of patterns most easily by identifying our problem with the classical problem of gambler's ruin. A pattern, read from right to left, may be interpreted as the record of a penny tossing game in which a gambler with initial capital $h = n - 2k$, playing against an adversary with unit initial capital, is ruined at the nth toss. The probability of such ruin is on the one hand $P(n, k)/2^n$; but on the other hand formulae for it are well known (see, for example, [1], p. 304, problem 6). In fact,

$$P(n, k) = (w_h + w_{3h+2} + w_{5h+4} + \cdots) - (w_{h+2} + w_{3h+4} + w_{5h+6} + \cdots), \tag{1}$$

where $h = n - 2k$, and

$$w_z = \frac{z}{n}\binom{n}{\frac{1}{2}(n - z)}$$

is the number of ways in which a gambler with initial capital z can be ruined at the nth toss when playing against an infinitely rich adversary.

If we take advantage once more of the restriction $k < n/3$, only two terms of (1) differ from zero, so that

$$P(n, k) = \frac{n - 2k}{n}\binom{n}{k} - \frac{n - 2k + 2}{n}\binom{n}{k - 1}. \tag{2}$$

Let $Q(n, k)$ denote the number of patterns with at most k minus signs uppermost. Summing (2) we obtain

$$Q(n, k) = \frac{n - 2k}{n}\binom{n}{k} = \binom{n - 1}{k} - \binom{n - 1}{k - 1}.$$

Recalling that there are $2n$ cycles for each pattern and 2^n cycles in all, while

under the null hypothesis these 2^n cycles are equally likely, we find

$$\Pr\{K \leqq k\} = (n - 2k)\binom{n}{k} \Big/ 2^{n-1}.$$

The table gives values of $\Pr\{K \leqq k\}$ to $5D$ for $n = 1(1)30$, and $k < n/3$.

Table of $\Pr\{K \leqq k\}$, for $k < n/3$.

	k									
n	0	1	2	3	4	5	6	7	8	9
1	1.00000									
2	1.00000									
3	.75000									
4	.50000	1.00000								
5	.31250	.93750								
6	.18750	.75000								
7	.10938	.54688	.98438							
8	.06250	.37500	.87500							
9	.03516	.24609	.70312							
10	.01953	.15625	.52734	.93750						
11	.01074	.09668	.37598	.80566						
12	.00586	.05859	.25781	.64453						
13	.00317	.03491	.17139	.48877	.87280					
14	.00171	.02051	.11108	.35547	.73315					
15	.00092	.01190	.07050	.24994	.58319					
16	.00049	.00684	.04395	.17090	.44434	.79980				
17	.00026	.00389	.02698	.11414	.32684	.66095				
18	.00014	.00220	.01634	.07471	.23346	.52295				
19	.00007	.00123	.00978	.04805	.16264	.39922	.72450			
20	.00004	.00069	.00580	.03044	.11089	.29572	.59143			
21	.00002	.00038	.00340	.01903	.07420	.21347	.46575			
22	.00001	.00021	.00198	.01175	.04883	.15068	.35578	.65057		
23	.00001	.00012	.00115	.00718	.03167	.10429	.26474	.52605		
24	0	.00006	.00066	.00434	.02027	.07094	.19254	.41259		
25		.00003	.00038	.00260	.01282	.04750	.13723	.31517	.58020	
26		.00002	.00021	.00155	.00802	.03137	.09606	.23525	.46559	
27		.00001	.00012	.00092	.00497	.02045	.06616	.17202	.36390	
28		.00001	.00007	.00054	.00305	.01318	.04491	.12350	.27789	.51490
29		0	.00004	.00031	.00186	.00841	.03008	.08722	.20786	.41055
30			.00002	.00018	.00112	.00531	.01991	.06067	.15263	.31987

Although $\Pr\{K < n/3\}$ tends to 0 as $n \to \infty$, it does not fall below 5 percent until $n = 72$, or below 1 percent until $n = 102$. If the test proves useful, it

may be desirable to consider the distribution of K for $k \geqq n/3$, where the results are likely to be less simple and neat.

REFERENCE

[1] WILLIAM FELLER, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1950.

ON THE CONVERGENCE OF EMPIRIC DISTRIBUTION FUNCTIONS[1]

BY J. R. BLUM

Indiana University

1. Summary. Let μ be a probability measure on the Borel sets of k-dimensional Euclidean space E_k. Let $\{X_n\}$, $n = 1, 2, \cdots$, be a sequence of k-dimensional independent random vectors, distributed according to μ. For each $n = 1, 2, \cdots$ let μ_n be the empiric distribution function corresponding to $X_1, \cdots, X_n$, i.e., for every Borel set $A \, \varepsilon \, E_k$, we define $\mu_n(A)$ to be the proportion of observations among $X_1, \cdots, X_n$ which fall in A.

Let $\mathfrak{A}$ be the class of Borel sets in E_k defined below. The object of this paper is to prove that $P\{\lim_{n\to\infty} \sup_{A\varepsilon\mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1$.

2. Introduction. Let $F(x)$ be a distribution function on the real line and let $\{X_n\}$, $n = 1, 2, \cdots$, be a sequence of independent random variables distributed according to F. For each $n = 1, 2, \cdots$ let $F_n(x)$ be the empiric distribution function corresponding to $X_1, \cdots, X_n$. The well-known theorem of Glivenko-Cantelli (see, e.g., Fréchet [1]) states that

$$P\{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\} = 1.$$

Fortet and Mourier [2] have proved several theorems on the convergence of empiric distribution functions in a separable metric space E. In particular, they show that if E is a Euclidean space and μ is a probability measure on E which is absolutely continuous with respect to Lebesgue measure, then

$$P\{\lim_{n\to\infty} \sup_{A\varepsilon\mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1, \tag{2.1}$$

where $\mathfrak{A}$ is the collection of open half-spaces in E. Wolfowitz [3] proved that (2.1) holds without any assumptions on μ. In this note we prove that if μ is absolutely continuous with respect to Lebesgue measure, then (2.1) holds for a considerably more general class of sets.

To avoid repetition we shall assume from now on that every set considered is

Received October 29, 1954.

[1] Work done with the support of the Office of Ordnance Research, U.S. Army.

a Borel subset of E_k. Let $\mathfrak{A}_1$ be the class of sets A each of which possesses the following property. If $x = (x_1, \cdots, x_k) \varepsilon A$ and $y = (y_1, \cdots, y_k)$ is such that $y_i < x_i$ for $i = 1, \cdots, k$, then $y \varepsilon A$. Let $\mathfrak{A}_j$, $j = 2, \cdots, 2^k$, be the $2^k - 1$ classes of sets which can be obtained by reversing, one at a time, the k inequalities occurring in the definition of $\mathfrak{A}_1$. Let $\mathfrak{A} = \bigcup_{j=1}^{2^k} \mathfrak{A}_j$. In this note we shall prove the following theorem.

Theorem. *If μ is absolutely continuous with respect to Lebesgue measure, then*

$$P\{\lim_{n\to\infty} \sup_{A \varepsilon \mathfrak{A}} |\mu_n(A) - \mu(A)| = 0\} = 1.$$

3. Proof of the theorem. In proving the theorem we shall restrict ourselves to the class $\mathfrak{A}_1$. The method of proof also applies to each of the classes $\mathfrak{A}_2, \cdots, \mathfrak{A}_{2^k}$, and consequently, from elementary considerations, the theorem holds for $\mathfrak{A}$. The method of proof depends on the following lemma.

Lemma 1. *Let $\mathfrak{B}$ be a class of sets and suppose for each $\rho > 0$ there exists a finite class of sets $\mathfrak{B}(\rho)$ such that for each $B \varepsilon \mathfrak{B}$ there exist sets B_1 and B_2 in $\mathfrak{B}(\rho)$ satisfying*

i) $$B_1 \subset B \subset B_2,$$

ii) $$\mu(B_2) - \mu(B_1) \leqq \rho.$$

Then $P\{\lim_{n\to\infty} \sup_{B \varepsilon \mathfrak{B}} |\mu_n(B) - \mu(B)| = 0\} = 1$.

The proof of the lemma is a direct consequence of the strong law of large numbers and is omitted here.

In proving the theorem we shall assume that $k = 2$. It will be clear from the sequel that the method of proof applies to arbitrary k, although the details become vastly more complicated.

Let R be a closed square in the plane which is subdivided into m^2 subsquares of equal area by dividing each side into m equal length intervals. Let $\mathfrak{A}_1(R)$ be the class of sets of the form $A \cap R$, with $A \varepsilon \mathfrak{A}_1$. For each set $T \varepsilon \mathfrak{A}_1(R)$ let $B(T)$ be the set of boundary points of T with the exception of those lying on the south and west boundaries of R. If $x = (x_1, x_2) \varepsilon R$, we shall say that x lies in a subsquare if it lies in the interior or on the north or east boundary of the subsquare. Let $N(T)$ be the number of distinct subsquares in which the points of $B(T)$ lie. Then we have the following lemma.

Lemma 2. *For every $T \varepsilon \mathfrak{A}_1(R)$, $N(T) \leqq 2m - 1$.*

Proof. We may assume that the coordinates of the corners of the subsquares are of the form (i, j) with $i = 0, \cdots, m$; $j = 0, \cdots, m$. Now consider the $2m - 1$ lines of the form $f(x) = x + k$, with $k = -m + 1, -m + 2, \cdots, m - 1$. By identifying each subsquare with the coordinates of its northeast corner it is easily seen that through each subsquare passes one and only one of these lines. Let $T \varepsilon \mathfrak{A}_1(R)$. We shall show that on every line of the form $f(x) = x + k$ there lies at most one point of $B(T)$. For suppose $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two distinct points of $B(T)$, and both lying on a line $f(x) = x + k$. Assume that $x_i < y_i$, $i = 1, 2$. Then we can find a point $z = (z_1, z_2) \varepsilon T$, with $x_i < z_i$, $i = 1, 2$. But this contradicts the fact that $x \varepsilon B(T)$. From this it follows that

each line $f(x) = x + k$ passes through at most one subsquare containing points of $B(T)$. Since there are $2m - 1$ such lines the lemma follows.

Let (i_1, j_1) be the coordinates of a corner of a subsquare with either $i_1 = 0$ or $j_1 = m$, and let (i_2, j_2) be the coordinates of another corner of a subsquare with either $i_2 = m$ or $j_2 = 0$. By a path P in R we shall mean a linear continuum of points connecting (i_1, j_1) with (i_2, j_2) and satisfying in addition:

i) Every point $x = (x_1, x_2) \varepsilon P$ lies on the boundary of a subsquare of R.

ii) If $x = (x_1, x_2) \varepsilon P$ and $y = (y_1, y_2) \varepsilon P$, and if $x_1 \leqq y_1$, then $x_2 \geqq y_2$.

By induction on m it is easily verified that there are at most finitely many paths P in R. To each path P we associate two sets $T_1(P)$ and $T_2(P)$ in $\mathfrak{A}_1(R)$ with $B(T_1) = B(T_2) = P$ and such that $T_1(P)$ contains all points of P and $T_2(P)$ contains no points of P. Let $\mathfrak{A}_{1,m}(R)$ be the class of all sets obtained in this manner for all possible paths P. Then $\mathfrak{A}_{1,m}(R)$ is clearly a finite class of sets for each integer m.

Let $T \varepsilon \mathfrak{A}_1(R)$, and let ρ be a positive number. For any positive integer m we may then choose two sets T_1 and T_2 in $\mathfrak{A}_{1,m}(R)$ such that $T_1 \subset T \subset T_2$, and such that if T' and T'' are in $\mathfrak{A}_{1,m}(R)$ and if $T_1 \subset T' \subset T \subset T'' \subset T_2$, then $T_1 = T'$ and $T_2 = T''$. From the choice of T_1 and T_2 it is clear that $T_2 - T_1$ is contained in the set of subsquares which contain $B(T)$. Let $L(U)$ be the Lebesque measure of a set U. Then from Lemma 2 it follows that $L(T_2 - T_1) \leqq L(R)\, N(T) / m^2 \leqq L(R)\,(2m - 1) / m^2$. Since μ is absolutely continuous with respect to L, we may choose an integer m such that $\mu(T_2 - T_1) < \rho$. Applying Lemma 1 we obtain the following lemma.

LEMMA 3. $P\{\lim_{n\to\infty} \sup_{T\varepsilon\mathfrak{A}_1(R)} |\mu_n(T) - \mu(T)| = 0\} = 1$.

Let ρ be a positive number. Let $A \varepsilon \mathfrak{A}_1$, and let R be a square with $\mu(R) > 1 - \rho/4$. Write $A = A_1 \cup A_2$, where $A_1 = A \cap R$, $A_2 = A \cap \bar{R}$, and where $\bar{R}$ is the complement of R. By virtue of Lemma 3 it suffices to show that $\lim_{n\to\infty} \sup_{A_2} |\mu_n(A_2) - \mu(A_2)| = 0$ on a set of sample sequences of probability one. Now consider any sample sequence in the set of probability one for which $\lim_{n\to\infty} \mu_n(\bar{R}) = \mu(\bar{R})$. Choose n so large that $\mu_n(\bar{R}) < \rho/2$. Since $0 \leqq \mu_n(A_2) \leqq \mu_n(\bar{R})$ and $0 \leqq \mu(A_2) \leqq \mu(\bar{R})$, we obtain $|\mu_n(A_2) - \mu(A_2)| < \rho/2$, uniformly in $A_2 \subset \bar{R}$, and the proof of the theorem is complete.

It appears to be a reasonable conjecture that the theorem is true without the condition of absolute continuity. One can easily construct examples which show that the method used in this note will no longer apply in the general situation. It would be of some interest to extend the result to the general case.

REFERENCES

[1] MAURICE FRÉCHET, "Généralités sur les Probabilités. Élements aléatoire", Gauthier-Villars, Paris, 1950.

[2] R. FORTET AND E. MOURIER, "Convergence de la répartition empirique vers la répartition théorique," Ann. Sci. Ecole Norm. Sup., (3) Vol. 70 (1953).

[3] J. WOLFOWITZ, "Generalization of the theorem of Glivenko-Cantelli," Ann. Math. Stat., Vol. 25, (1954), pp. 131–138.

CALCULATION OF EXACT SAMPLING DISTRIBUTION OF RANGES FROM A DISCRETE POPULATION[1]

By Irving W. Burr

Purdue University

1. Introduction. The exact sampling distribution for ranges is known for but few populations, and general information on moments of the range is incomplete. This note gives a method for calculating the exact sampling distribution for discrete universes having a finite range and approximating those for populations with an infinite range.

2. Derivation. Consider a random variable X defined on integers a to b, both finite. Let p_i be the probability that X is i, and $p(R)$ be the probability that the range takes the value R. Then for a sample of n X's from the population (drawn with replacement) we have

$$p(R) = \sum_{i=a}^{b-R} \sum_{r=1}^{n-1} \sum_{s=1}^{n-r} \frac{n!\, p_i^r p_{i+R}^s}{r!\, s!\, (n-r-s)!} (p_{i+1} + \cdots + p_{i+R-1})^{n-r-s}, \tag{1}$$

since the summand contains at least one X at i and at least one X at $i + R$ and those X's not at these values are all between, and the summation is over all possible such samples. To obtain a more useful form we let

$$M(i, R) = \sum_{j=1}^{i+R} p_j. \tag{2}$$

Then

$$p(R) = \sum_{i=a}^{b-R} \sum_{r=1}^{n-1} \sum_{s=1}^{n-r} \frac{n!\, p_i^r p_{i+R}^s}{r!\, s!\, (n-r-s)!} M^{n-r-s}(i+1, R-2)$$

$$= \sum_{i=a}^{b-R} [\text{terms of } M^n(i, R) \text{ containing at least one } i \text{ and at least one } i + R].$$

To get the desired terms of $M^n(i, R)$, we first subtract from it all of those terms which fail to contain any $i + R$, namely, $M^n(i, R - 1)$. Then we also subtract off those which fail to contain any i, namely $M^n(i + 1, R - 1)$. But these two expressions overlap to the extent of $M^n(i + 1, R - 2)$, that is, terms with neither i nor $i + R$. So this must be added back on. Thus we have

$$\begin{aligned} p(R) = \sum_{i=a}^{b-R} [M^n(i, R) &- M^n(i, R-1) \\ &- M^n(i+1, R-1) + M^n(i+1, R-2)]. \end{aligned} \tag{3}$$

Received November 27, 1953; revised July 29, 1954.

[1] Presented by title to the Institute, December 27, 1951, at Boston.

To systematize calculation, another form is desirable. Let

$$C_R = \sum_{i=a+1}^{b-R-1} M^n(i, R), \tag{4}$$

$$E_R = M^n(a, R) + M^n(b - R, R). \tag{5}$$

Then we have

$$p(R) = C_R + E_R - 2C_{R-1} - E_{R-1} + C_{R-2}. \tag{6}$$

Formulas (3) and (6) are appropriately modified for $R = 0, 1, b - a - 1$, and $b - a$.

3. Calculation. In computing the $p(R)$, the universe probabilities can best be listed as integer frequencies, as small as possible. Then sums of consecutive frequencies, two at a time, three at a time, etc., are formed, the resulting table being of the same form as a table of differences. Then the C_k and E_k are found by forming sums of nth powers of these table entries. The appropriate modifications of (6) are made by omitting terms naturally absent from this table.

4. An Example. Formula (6) enables us to study the effect on ranges of non-normality in the population. Thus we may compare the following two distributions: One a discrete distribution with probabilities approximately proportional to normal curve areas and the other approximately proportional to those of a well-skewed Pearson Type III.

X	0	1	2	3	4	5	6	7	8	9	10	11
f_1	.005	.015	.050	.115	.195	.240	.195	.115	.050	.015	.005	.000
f_2	.01	.13	.22	.21	.17	.11	.07	.04	.02	.01	.00	.01

The respective characteristics are

$$\mu = 5.00 \qquad \sigma = 1.71 \qquad \alpha_3 = 0 \qquad \alpha_4 = 3.02$$

$$\mu = 3.45 \qquad \sigma = 1.99 \qquad \alpha_3 = .99 \qquad \alpha_4 = 4.21$$

The respective distributions of range $n = 5$ are the following:

R	0	1	2	3	4	5	6	7	8	9	10	11
$p_1(R)$	.001	.031	.146	.239	.251	.179	.096	.040	.013	.003	.0005	
$p_2(R)$	.001	.028	.114	.203	.221	.180	.117	.063	.030	.020	.020	.002

The characteristics are respectively

$$\mu_R = 3.93 \qquad \sigma_R = 1.53 \qquad \alpha_3 = .41 \qquad \alpha_4 = 3.01$$

$$\mu_R = 4.44 \qquad \sigma_R = 1.94 \qquad \alpha_4 = .73 \qquad \alpha_4 = 3.47$$

It can be seen that there is much less difference in skewness in the distributions

of R than in the original populations. The R distributions are in fact quite similar if allowance is made for the difference in population standard deviations. Hence we can have quite a bit of confidence in using normal curve constants when making control charts for ranges for moderately skewed populations and small sample sizes.

THE STOCHASTIC CONVERGENCE OF A FUNCTION OF SAMPLE SUCCESSIVE DIFFERENCES[1]

By Lionel Weiss

University of Virginia

1. Summary and introduction. Let $f(x)$ be a bounded density function over the finite interval $[A, B]$ with at most a finite number of discontinuities. Let $X_1, X_2, \cdots, X_n$ be independent chance variables each with the density $f(x)$. Define $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_n$ as the ordered values of $X_1, X_2, \cdots, X_n$, and T_i as $Y_{i+1} - Y_i$. Also define $R_n(t)$ as the proportion of the variates $T_1, \cdots, T_{n-1}$ not greater than $t/(n-1)$. We shall denote $[1 - \int_A^B f(x)e^{-tf(x)}\,dx]$ by $S(t)$, and $\sup_{t \geqq 0} |R_n(t) - S(t)|$ by $V(n)$. Then it is shown that as n increases, $V(n)$ converges stochastically to zero. The relation of this result to other results is discussed.

2. Proof of the stochastic convergence of $V(n)$ to zero.

Lemma 1. *If for each given t, $R_n(t)$ converges stochastically to $S(t)$ as n increases, then $V(n)$ converges stochastically to zero.*

Proof. We must show that for any given positive numbers ϵ and δ, there is a positive integer $N(\epsilon, \delta)$ such that if $n > N(\epsilon, \delta)$, then $P[V(n) < \epsilon] > 1 - \delta$. We can find a finite set of values $t_0 < t_1 < \cdots < t_s$ such that

$$S(t_0) < \tfrac{1}{2}\epsilon, \qquad 1 - S(t_s) < \tfrac{1}{2}\epsilon, \qquad S(t_{i+1}) - S(t_i) < \tfrac{1}{2}\epsilon,$$
$$i = 0, 1, \cdots, s - 1.$$

Also, by the hypothesis of the lemma and other familiar considerations, we can find a positive integer, say $N(\epsilon, \delta)$, such that if $n > N(\epsilon, \delta)$,

$$P[|R_n(t_i) - S(t_i)| < \tfrac{1}{2}\epsilon \quad \text{for} \quad i = 0, \cdots, s] > 1 - \delta.$$

But then the lemma is proved, for it is easily verified that if $|R_n(t_i) - S(t_i)| < \frac{1}{2}\epsilon$ simultaneously for $i = 0, \cdots, s$, then $|R_n(t) - S(t)| < \epsilon$ simultaneously for all $t \geqq 0$.

Lemma 2. *Let $X_1, \cdots, X_n$ be independent chance variables each with a uniform distribution on* $[0, 1]$. *Let M denote the number of these variables falling in the closed*

Received August 6, 1954.

[1] Research under a grant from the Institute for Research in the Social Sciences, University of Virginia.

interval $[C, D]$, *where* $0 \leqq C < D \leqq 1$, *and let* $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_M$ *denote the ordered values of the variables in* $[C, D]$. *Define* $W_0 = Y_1 - A$, *and* $W_i = Y_{i+1} - Y_i$ *for* $i = 1, \cdots, M - 1$. *Finally, define* $L(n, t)$ *as the total number of values of* $W_1, \cdots, W_{M-1}$ *which are not greater than* $t / (n - 1)$ *for a given* $t \geqq 0$. *Then* $L(n, t) / (n - 1)$ *converges stochastically to* $(D - C)[1 - e^{-t}]$ *as n increases.*

PROOF. We denote $(D - C)$ by G, and by $K(n, t)$ the total number of $W_0, \cdots, W_{M-1}$ not greater than $t / (n - 1)$. Clearly, the lemma will be proved if we can show that $K(n, t) / (n - 1)$ converges stochastically to $G(1 - e^{-t})$ as n increases. The distribution of M is binomial, with parameters G and n, and the joint conditional density of $Y_1, \cdots, Y_M$ given M is $M! / G^M$ in the region $C \leqq Y_1 \leqq \cdots \leqq Y_M \leqq D$, and zero elsewhere. Thus the joint conditional density of $W_0, \cdots, W_{M-1}$ given M is $M! / G^M$ in the region $W_i \geqq 0$ and $\sum_{i=0}^{M-1} W_i \leqq G$.

Define Z_i to be 1 if $W_i \leqq t / (n - 1)$, and zero otherwise. By the symmetry of the joint conditional distribution of $W_0, \cdots, W_{M-1}$, $E\,K(n, t) = E\{M \cdot E[Z_0 \mid M]\}$. The conditional density of W_0 given M is $M(G - w)^{M-1} / G^M$ for $0 \leqq w \leqq G$. Thus $E[Z_0 \mid M] = 1 - (1 - t / (n - 1)G)^M$, assuming $G \geqq t / (n - 1)$, which involves no loss of generality, since G is fixed and we are interested in what happens as n increases. By routine manipulations of the moment generating function of M, we find that

$$EK(n, t) = E\{M[1 - (1 - t / (n - 1)G)^M]\}$$
$$= nG - [1 - t / (n - 1)]^{n-1}[nG - nt / (n - 1)],$$

From this, we find that $E[K(n, t) / (n - 1)]$ approaches $G(1 - e^{-t})$ as n increases. Next we examine

$$E\left[\frac{K(n,t)}{n-1}\right]^2 = \left(\frac{1}{n-1}\right)^2 E\left(\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} Z_i Z_j\right)$$
$$= \left(\frac{1}{n-1}\right)^2 E\left(\sum_{0}^{M-1} Z_i^2\right) + \left(\frac{1}{n-1}\right)^2 E\left(\sum_{i \neq j}\sum Z_i Z_j\right).$$

But $\sum Z_i^2 = \sum Z_i$, and from above we have that $E[\sum Z_i / (n - 1)]$ approaches $G(1 - e^{-t})$ an n increases. Therefore $E[K(n, t) / (n - 1)]^2$ has the same limit as

$$\left(\frac{1}{n-1}\right)^2 E\left[\sum_{i \neq j}\sum Z_i Z_j\right] = \left(\frac{1}{n-1}\right)^2 E[M(M - 1) \cdot E(Z_0 Z_1 \mid M)].$$

This last equality holds because of the symmetry of the distribution of $W_0, \cdots, W_{M-1}$. The joint conditional density of W_0, W_1 given M is $M(M - 1)(G - w_0 - w_1)^{M-2} / G^M$ for $w_0, w_1 \geqq 0$ and $w_0 + w_1 \leqq G$. Thus

$$E(Z_0 Z_1 \mid M) = 1 - 2\left(1 - \frac{t}{(n-1)G}\right)^M + \left(1 - \frac{2t}{(n-1)G}\right)^M,$$

provided $G \geqq 2t / (n-1)$, which involves no loss of generality. Therefore $E[K(n, t) / (n-1)]^2$ has the same limit as

$$\left(\frac{1}{n-1}\right)^2 E\left\{M(M-1)\left[1 - 2\left(1 - \frac{t}{(n-1)G}\right)^M + \left(1 - \frac{2t}{(n-1)G}\right)^M\right]\right\}$$

$$= \left[\frac{nG^2}{n-1}\right]\left[1 - 2\left(1 - \frac{t}{(n-1)G}\right)^2\left(1 - \frac{t}{(n-1)}\right)^{n-2} + \left(1 - \frac{2t}{(n-1)G}\right)^2\left(1 - \frac{2t}{(n-1)}\right)^{n-2}\right].$$

This last expression approaches $G^2[1 - 2e^{-t} + e^{-2t}] = [G(1 - e^{-t})]^2$ as n increases. But this proves Lemma 2, since the variance of $K(n, t) / (n-1)$ approaches zero as n increases.

With a few simple changes in notation, Lemma 1 serves to show that $\sup_{t \geq 0} |L(n, t) / (n-1) - G(1 - e^{-t})|$ converges stochastically to zero as n increases. Also, when $G = 1$, Lemmas 1 and 2 prove that $V(n)$ converges stochastically to zero for the special case where $f(x) = 1$ on the interval [0, 1].

Now we turn to the proof that $V(n)$ converges to zero in the general case. We denote $\int_A^x f(x)\, dx$ by $F(x)$. By the assumptions about $f(x)$ listed in Section 1, given any positive number γ, we can break the interval $[A, B]$ into a finite number $k(\gamma)$ of subintervals $(a_0, a_1), \cdots, (a_{k(\gamma)-1}, a_{k(\gamma)})$, with $a_0 = A$ and $a_{k(\gamma)} = B$, such that in the interior of each subinterval (a_i, a_{i+1}), $f(x)$ is continuous, and for any x in the subinterval, $|f(x) - f(a_i)| < \gamma$ for $i = 0, \cdots, k(\gamma) - 1$. Choose any particular subinterval (a_i, a_{i+1}), and let $Q_1 \leqq Q_2 \leqq \cdots \leqq Q_M$ denote the ordered values of those variables $X_1, \cdots, X_n$ which fall in (a_i, a_{i+1}), while T'_j shall denote $Q_{j+1} - Q_j$ for $j = 1, \cdots, M-1$. Denote $F(Q_{j+1}) - F(Q_j)$ by W_j. Then, defining $L_i(n, t)$ in terms of $W_1, \cdots, W_{M-1}$ as in Lemma 2, with the subscript i to show that we are dealing with the interval (a_i, a_{i+1}), we have that $\sup_{t \geq 0} | L_i(n, t)/(n-1) - \{F(a_{i+1}) - F(a_i)\}(1 - e^{-t}) |$ converges stochastically to zero as n increases. (This is so because $F(X_j)$ has the rectangular distribution over (0, 1) for any j.) By construction, we have

$$W_j = F(Q_{j+1}) - F(Q_j) = T'_j(f(a_i) + \theta_j), \qquad |\theta_j| < \gamma.$$

Therefore, if $T'_j \leqq t/(n-1)$, then $W_j \leqq (f(a_i) + \gamma)t/(n-1)$, and conversely. Then, letting $K_i(n, t)$ denote the number of the values T'_j which are not greater than $t/(n-1)$, we have

$$L_i(n, t(f(a_i) - \gamma)) \leqq K_i(n, t) \leqq L_i(n, t(f(a_i) + \gamma)).$$

Using $R_n(t)$ as defined in Section 1, this becomes

$$\frac{1}{n-1} \sum_{i=0}^{k(\gamma)-1} L_i(n, t(f(a_i) - \gamma)) - k(\gamma) \leqq R_n(t)$$

$$\leqq \frac{1}{n-1} \sum_{i=0}^{k(\gamma)-1} L_i(n, t(f(a_i) + \gamma)) + k(\gamma).$$

Given any positive values ϵ, δ, we first choose γ so small that

$$\left| \sum_{i=0}^{k(\gamma)-1} (F(a_{i+1}) - F(a_i))e^{-tf(a_i)}e^{t\bar{\gamma}} - \int_A^B f(x)e^{-tf(x)}\,dx \right| < \tfrac{1}{2}\varepsilon$$

for $\bar{\gamma}$ equal to either γ or $-\gamma$. Then we choose $N(\epsilon, \delta)$ so large that if $n > N(\epsilon, \delta)$, then $k(\gamma)/(n-1) < \frac{1}{4}\epsilon$, and also

$$P[\sup_{t \geq 0} | L_i(n, t)/(n-1) - \{F(a_{i+1}) - F(a_i)\}(1 - e^{-t}) | < \epsilon/4k(\gamma),$$
$$i = 0, \cdots, k(\gamma) - 1] > 1 - \delta.$$

But then if $n > N(\epsilon, \delta)$,

$$P\begin{bmatrix} \sum_{i=0}^{k(\gamma)-1} \{F(a_{i+1}) - F(a_i)\}(1 - e^{-tf(a_i)}e^{t\gamma}) - \frac{1}{2}\epsilon \leq R_n(t) \\ \leq \sum_{i=0}^{k(\gamma)-1} \{F(a_{i+1}) - F(a_i)\}(1 - e^{-tf(a_i)}e^{-t\gamma}) + \frac{1}{2}\epsilon \end{bmatrix} > 1 - \delta,$$

or $P\{ | R_n(t) - S(t) | < \epsilon \} > 1 - \delta$. This shows that for any given t, $R_n(t)$ converges stochastically to $S(t)$. Then by Lemma 1, $V(n)$ converges stochastically to zero.

The same results hold with only slight modifications in the argument when $A = -\infty$ and/or $B = \infty$, provided that there exist finite numbers A' and B', with $A' < B'$, such that $f(x)$ is nondecreasing in the interval $(-\infty, A')$ and is nonincreasing in the interval (B', ∞).

3. Relation to other results. The stochastic convergence of certain functions of $T_1, \cdots, T_{n-1}$ can be proved simply by the use of these results. For example, Sherman [1] studied the chance variable Ω_n defined as

$$\tfrac{1}{2}\sum_{i=1}^{n-1} \left| T_i - \frac{1}{n+1} \right| + \tfrac{1}{2}\left| Y_1 - A \right| + \tfrac{1}{2}\left| B - Y_n \right|.$$

Let us assume that A is the least upper bound of all numbers a such that $F(a) = 0$, and B is the greatest lower bound of all numbers b such that $F(b) = 1$. Then as n increases, $| Y_1 - A |$ and $| B - Y_n |$ converge stochastically to zero. Thus Ω_n converges stochastically to a constant if and only if $U_n = \frac{1}{2}\sum_{i=1}^{n-1} | T_i - 1/(n-1) |$ converges stochastically to the same constant. We can write U_n as $S_n + V_n$ where

$$S_n = \tfrac{1}{2} \sum_{i:\, T_i \leq 1/(n-1)} \left\{ \frac{1}{n-1} - T_i \right\}, \qquad V_n = \tfrac{1}{2} \sum_{i:\, T_i > 1/(n-1)} \left\{ T_i - \frac{1}{n-1} \right\}.$$

But $\frac{1}{2}\sum_{i=1}^{n-1}\{T_i - 1/(n-1)\} = V_n - S_n$ converges stochastically to $\frac{1}{2}(B - A - 1)$. Thus U_n converges stochastically to a constant if and only if S_n converges stochastically to a constant. We can write

$$S_n = \frac{n-1}{2} \int_0^1 \left(\frac{1}{n-1} - \frac{t}{n-1} \right) dR_n(t) = \tfrac{1}{2}\left[R_n(1) - \int_0^1 t\, dR_n(t) \right].$$

Integrating by parts, we find that $S_n = \frac{1}{2}\int_0^1 R_n(t)\, dt$. By the result proved in Section 2 this last expression converges stochastically to

$$\frac{1}{2}\int_0^1 \left[1 - \int_A^B f(x)\, e^{-tf(x)}\, dx\right] dt = \frac{1}{2}\left[1 + \int_A^B e^{-f(x)}\, dx - (B - A)\right].$$

Therefore Ω_n converges stochastically to $\frac{1}{2}(1 + A - B) + \int_A^B e^{-f(x)}\, dx$. For the special case $A = 0$ and $B = 1$, this is essentially the result contained in theorems 3 and 4 of [1].

REFERENCE

[1] B. Sherman, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

Note added in proof. Professor Julius Blum has pointed out that Lemma 2 holds with the words "converges stochastically" replaced by "converges with probability one." Then it is easily seen that all the results above hold when this replacement is made.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Chapel Hill meeting of the Institute, April 22–23, 1955)

1. Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples. Part I. The Normal Distribution up to Samples of Size 10. A. E. Sarhan and B. G. Greenberg, University of North Carolina.

The variances and covariances of the order statistics for samples of sizes $\leqq 20$ from a normal distribution were calculated to 10 decimal places from Teichroew's tables of the expected value of the product of two order statistics. By the use of these values, and with the table of expected values of Rosser, the best linear estimates of the mean and standard deviation were calculated from singly and doubly censored samples up to samples of size 10. This was accomplished by applying the method of least squares to the linear combination of the ordered known observations to obtain unbiased estimates with minimum variance. The variances of the estimates were also calculated. An alternative linear estimate was derived for larger values of n which can be used to obtain estimates from doubly censored samples.

2. An Application of Chung's Lemma to the Kiefer-Wolfowitz Stochastic Approximation Procedure. Cyrus Derman, Syracuse University.

Let $M(x)$ be a strictly increasing regression function for $x < \theta$, and strictly decreasing regression function for $x > \theta$. Kiefer and Wolfowitz (Ann. Math. Stat., Vol. 23 (1952), pp. 462–466) suggested a recursive scheme for estimating θ. They proved, under certain regularity conditions, that their scheme converges stochastically to θ. Their conditions

exclude the case $M(x) = K - K'(x - \theta)^2$ where K and K' are constants ($K' > 0$). Conditions, which do not exclude the above case, are given here for their scheme to converge stochastically to θ. Under stronger conditions (the above case still not excluded) convergence to the normal distribution is proved. The main tool used in the analysis is a lemma due to Chung (Ann. Math. Stat., Vol. 25 (1954), pp. 463–483).

3. Simplified Estimators Based on Order Statistics. (Preliminary Report.) BENJAMIN EPSTEIN, Wayne University.

Best linear unbiased estimates based on order statistics have been given recently by A. E. Sarhan [*Ann. Math. Stat.*, Vol. 25 (1954), pp. 317–328] for the mean and standard deviation of a number of distributions. It is assumed in that paper that all observations are known. In a paper given at the Berkeley meeting in December, 1954, Sarhan considered the same estimation problem in the case where some of the ordered observations may be missing. Here we give unbiased estimators which are much simpler in the sense that they can be expressed in terms of only a few of the order statistics about which we have information. Efficiency of the suggested estimators is high for small sample sizes.

4. Distribution of the Difference Between the Two Largest Sample Values. (Preliminary Report.) A. ZINGER and J. ST-PIERRE, University of Montreal.

A decision procedure to select the population with the largest mean, proposed by R. C. Bose and J. St.-Pierre (Ann. Math. Stat., Vol. 25 (1954), p. 813), involves the auxiliary statistic $y = x_{(0)} - x_{(1)}$, where $x_{(0)}$ and $x_{(1)}$ are respectively the largest and second largest values in a sample of $n + 1$ variates. The distribution of this statistic is obtained with a method simpler than the one already used by the senior author. The final result involves iterated integrals of the normal density over simple limits. The general form can be easily reduced to neat expressions in the case of lower dimensions. The densities of 3, 4, and 5 dimensions have been extensively tabulated and a recursion formula established between the densities. The establishment of a recursion formula in the general case is being worked on.

5. Some Continuous Monte Carlo Methods for the Dirichlet Problem. MERVIN E. MULLER, Cornell University.

Monte Carlo techniques are introduced using stochastic models which are Markov processes. This material includes the N-dimensional spherical, general spherical, and general Dirichlet domain processes. These processes are proved to converge with probably 1 and thus yield direct statistical estimates of the solution to the N-dimensional Dirichlet problem. The results are obtained without requiring any further restrictions on the boundary or the function defined on the boundary in addition to those required for the existence and uniqueness of the solution to the Dirichlet problem. A detailed study is made for the N-dimensional spherical process. This includes a study of the order of the average number of steps required for convergence. Asymptotic confidence intervals are obtained. When computing effort is measured in terms of the order of the average number of steps required for convergence, the often made conjecture that the computing effort of a Monte Carlo procedure should be a linear function of the dimensionality of the problem is shown to be true for the cases considered. Comments are included regarding the application of these processes on digital computers. Truncation methods are suggested.

6. On the Distribution of the Number of Successes in Independent Trials. Wassily Hoeffding, University of North Carolina.

Let S be the number of successes in n independent trials. Let p_i be the probability of success in the ith trial. The problem is considered of finding the maximum or the minimum of the expected value of a function of S when $E(S) = np$ is fixed, $0 < p < 1$. It is well known that the variance of S attains its maximum when the p_i are all equal. It is shown: (i) for any two integers b, c such that $0 \leqq b \leqq np \leqq c \leqq n$ the probability $P(b \leqq S \leqq c)$ attains its minimum if and only if all the p_i are equal, unless $b = 0$, $c = n$; (ii) for any strictly convex function g the expected value $Eg(S)$ attains its maximum if and only if all the p_i are equal. The maximum and the minimum of $P(S \leqq c)$, $0 \leqq c \leqq n$, are determined. These results are obtained with the aid of some theorems concerning the extrema of $Eg(S)$, where g is an arbitrary function. For example, the maximum and the minimum of $Eg(S)$ are attained at points $(p_1, \cdots, p_n)$ whose coordinates take on at most three different values, only one of which is distinct from 0 and 1. Statistical applications of (i) and (ii) are pointed out.

7. On the Solution of Truncated and Censored Sample Estimating Equations for Normal Populations. A. C. Cohen, Jr., University of Georgia.

To obtain maximum likelihood estimates of the mean and standard deviation of a normally distributed population from doubly truncated and from doubly censored samples, it is necessary to carry out the simultaneous solution of a pair of rather complicated nonlinear estimating equations. In this paper, iterative techniques for solving these equations are examined, and a procedure is developed which yields solutions of specified accuracy with less computational effort than required by other methods previously employed. A chart has been prepared which, for doubly truncated samples, permits a quick graphic solution to a degree of accuracy that is adequate for many purposes, and which provides a good first approximation for subsequent improvement through iteration when greater accuracy is demanded. A chart has also been devised to permit a quick graphic solution in the case of singly censored samples.

8. The Modified Mean Square Successive Difference and Related Statistics. Seymour Geisser, University of North Carolina.

In estimating the variance of a normal population, one uses the sample variance because of its optimum properties. In certain cases where there is an indeterminable trend in the data, it has been thought useful to estimate the variance by another statistic, namely, the mean square successive difference, the mean of the squared first differences, which under certain conditions, eliminates a good deal of the trend and is less biased than the sample variance. An explicit form of the exact distribution of this statistic seems, at least for the present, too difficult to obtain. However, by applying the device of Durbin and Watson, that is, by dropping from the mean square successive difference the middle term for an even number of observations and the two middle terms for the odd case, it is found that the quadratic form has double roots, thus making it possible to obtain the exact distribution in terms of elementary functions. In addition, one defines analogues of the Student t and the Fisher F using similarly modified statistics and proceeds to derive their exact distributions when the observations are independent and in a specific dependency case which has several properties in common with the stationary Markov process.

9. The Distribution of the Ratios of Certain Quadratic Forms in Time Series. (By Title.) SEYMOUR GEISSER, University of North Carolina.

In testing the hypothesis that successive members of a series of observations are serially correlated, a number of statistics have been proposed by various statisticians. R. L. Anderson gave the first exact distribution of a serial correlation coefficient using a circular definition. J. Durbin and G. Watson gave the exact distributions of several other statistics using double root methods. In this paper the work of Durbin and Watson has been extended for a non-null case of one of their statistics. Also, by introducing a new model, the exact distribution of a modified form of the von Neumann ratio has been derived in the non-null case. It has also been shown that this ratio provides a "best" test for the parameter involved.

10. The "Inefficiency" of the Sample Median for many Familiar Symmetric Distributions. J. T. CHU, University of North Carolina.

If the pdf of a certain distribution is symmetric and has an absolute maximum at the point of symmetry, a lower bound for var $\tilde{x}$, the variance of the sample median $\tilde{x}$ of a sample of size $2n + 1$ is $(2n + 1)/(2n + 3)$ multiplied by the variance of the asymptotic distribution of $\tilde{x}$ (which is normal). Therefore if sample size is not too small, the asymptotic variance of $\tilde{x}$ is for all practical purposes a lower bound for var $\tilde{x}$. If $\tilde{x}$ is asymptotically less efficient than $\bar{x}$, it is probable that $\tilde{x}$ is less efficient than $\bar{x}$ for most finite samples as well. For many symmetric distributions familiar to statisticians, such as triangular, Student's t, symmetric β, and Cauchy type distributions $(f(x) = C_\alpha/(1 + |x|^\alpha), -\infty < x < \infty, \alpha \geqq 4.65)$, not counting normal and rectangular distributions, it is shown that $\tilde{x}$ is for most sample sizes less efficient than $\bar{x}$.

11. On Some Stochastic Models of Behavioral Interaction of Organization Theory. DAVID ROSENBLATT, American University.

This paper treats certain stochastic models of behavioral interaction which constitute applications of a general approach to a calculus of behavior. Participants or groups in organizations are viewed as entities provided with a stochastic preference (or threshold) apparatus; entities engage in adaptive or reactive behavior by adjustment of the stochastic processes governing their own activities. Transition probabilities are modified in accord with experience of "relative success" of the entities in accord with certain criteria, e.g., organizational norms or observed actions of other entities. Modes of behavior are sequentially reinforced or inhibited as a result of the moves of entities in the course of interaction. Various types of memory structure are explicitly introduced. The "performance characteristics" of a given structure of interaction may be summarized by the expected individual and joint probability distributions of behavioral activity of each entity at each interaction transaction γ_k $(k = 1, 2, \cdots)$. Algorithms are developed for the determination of these "performance characteristics." For certain parametric characterizations (r entities, n_j decision alternatives and m_j preference valuation alternatives for the jth entity, $j = 1, 2, \cdots, r$), the algorithms lead to closed-form expressions. In the simplest cases, these become systems of linear difference equations. Many of the asymptotic results of stochastic learning theory may be readily obtained by specialization of the present models. (Work supported by the Office of Naval Research.)

12. On Inverting a Class of Patterned Matrices, Part I. S. N. ROY and A. E SARHAN, University of North Carolina.

In this note, inverses are given of a class of patterned matrices that occur in different sectors of statistics, e.g., least squares solutions relating to problems of estimation of

population parameters by ordered or unordered observations, analysis of variance and covariance, response surfaces, etc. The actual examples given here are illustrative and will be followed up later by other examples. In the technique given here of obtaining such inverses, use is made of the fact that (i) a non-singular square matrix has a unique inverse and (ii) for the class of patterned matrices considered it is possible to guess a form for the inverse with a few unknown (and thus flexible) parameters which could then be determined by equating to the identity matrix the product of the original matrix and the inverse that is guessed. At the moment the guess is just intuitive, but the authors believe there is a deeper calculus behind the whole thing, which may emerge later and thus make the inversion of such matrices an entirely trivial problem.

13. Convergence Properties of a General Stochastic Approximation Process. (Preliminary Report.) DONALD BURKHOLDER, University of North Carolina.

THEOREM. *Let* $\{R_n\}$ *be a sequence of Borel measurable functions,* $\theta, \sigma^2, c, d, x_1$ *real numbers,* Q *a function from the positive numbers to the natural numbers, and* $\{a_n\}$ *a positive number sequence such that:* (i) *for each natural number* n *and each real number* x *there is a random variable* $Z_n(x)$ *such that* $EZ_n(x) = R_n(x)$, $\text{Var}\,[Z_n(x)] \leq \sigma^2$, *and* $|R_n(x)| \leq c + d\,|x|$; (ii) if $0 < \epsilon < \infty$ *and* $0 < \delta_1 < \delta_2 < \infty$, *then* $(x - \theta)\, R_n(x) > 0$ *for* $|x - \theta| > \epsilon$, $n > Q(\epsilon)$, *and* $\Sigma a_n\, [\inf_{\delta_1 \leq |x-\theta| \leq \delta_2} |R_n(x)|] = \infty$; (iii) $\Sigma a_n^2 < \infty$. *Then the sequence* $\{x_n\}$ *of random variables defined recursively by* $x_{n+1} = x_n - a_n Z_n(x_n)$ *converges to* θ *with probability one.* The proof involves methods similar to those used by Blum, *Ann. Math. Stat.*, Vol. 25 (1954), pp. 382–386. Some immediate corollaries to the theorem are: (1) The Robbins-Munro process converges with probability one (Blum's Theorem 1). (2) The Kiefer-Wolfowitz process converges with probability one (under conditions less restrictive than those heretofore published; for instance, a regression function M where $M(x) = e^{-x^2}$ or $M(x) = -x^2$ is permissible). (3) There exists a strongly consistent sequence of estimates of the mode of a density function under fairly general conditions. (4) There exists a strongly consistent sequence of estimates of a root of a regression equation even when the variances around the regression line may not exist. The theorem, and hence also each of the corollaries (1), (2), and (3), has been generalized to the case where the number θ does not exist uniquely. This permits, for instance, the use of the Robbins-Munro process in the problem of estimating a quantile of a distribution function when the quantile is not unique.

14. Distribution of Rounding Off Errors in Some Numerical Processes, Part I. A. E. SARHAN, University of North Carolina.

The distributions of rounding-off errors in a product, quotient, raising to a power process, several combinations of these processes, and other special cases are derived. The moment generating functions and the first four moments are calculated. Expressions for the significance points at a given level are provided.

15. On a measure of the Information Provided by an Experiment. (Preliminary Report.) D. V. LINDLEY, University of Cambridge and University of Chicago.

An experiment consists in the observation of a random variable x with probability density $f(x \mid \theta)$, where θ is an unknown parameter. Let $p(\theta)$ be the probability density of θ, expressing the knowledge of θ prior to performing the experiment. Then the average amount of information provided by the experiment is defined to be $\int\int f(x \mid \theta) p(\theta) \log \{f(x \mid \theta)/p(x)\}$

$\cdot dx\, d\theta$, where $p(x) = \int f(x \mid \theta)p(\theta)\, d\theta$. This definition is suggested by the corresponding definition of Shannon's in connection with the rate of transmission of information in communication engineering. It is shown that it is always nonnegative, is not reduced by consideration of sufficient statistics alone, and if x and y are independent random variables for each θ, then the experiment in which y is observed is more informative if carried out before x is observed than if carried out after x has been observed. The definition enables comparisons to be made of different experiments but these comparisons are unlike those considered by Blackwell in that the losses in pursuing various courses of action are not considered. The ideas are therefore more relevant to the inference problem than the decision problem. Examples of the use of the definition, in particular for multivariate normal densities, are considered.

16. A Comparison Between Alternative Techniques Using Supplementary Information in Sample Survey Design. (Preliminary Report.) EL MAHDY SAID, North Carolina State College.

Three alternative methods of incorporating the advance information available on a variable X in the design of finite population sample surveys to estimate aggregate or mean values for a variable Y are studied. For a given sample size n and ignoring cost, the systems are (a) stratification with $s = n/2$ strata, (b) sampling without replacement with unequal selection probabilities such that the probability of including units $u_i u_j$ together in the sample is $P(u_i u_j) = n(n-1)X_i X_j(1/T_x - X_i + 1/T_x - X_j)/2T_x$ where $T_x = \sum_{i=1}^{N} X_i$ and the estimator used is $\bar{y}_p = \sum_{i=1}^{n} y_i/P(u_i)$, (c) ratio estimate. Formulas for the mean square errors of the estimators are derived for both linear and curvilinear relationships between X and Y. Exact comparison for $n = 2$, using discrete counterparts of some Pearson type III distributions for X, showed that (b) is superior to (c) except for ρ_{xy} very close to 1. Approximate comparisons were obtained for $n > 2$ assuming large N and continuous type III distributions. Variance with stratification was approximated by using the uniform distribution for X within the first $(s - 1)$ strata. Method (a) was found to be superior to (b) and (c) except when c_y (the coefficient of variation of Y) is in the neighbourhood of c_x. In certain instances the issue depends on ρ_{xy} alone; in others, on combination of ρ_{xy} and c_y.

17. The Canonical Distribution of the Non-central Rectangular Co-ordinates. MISS ALEYAMMA GEORGE, University of North Carolina and University of Travancore.

This paper is concerned with a matrix method of (a) deriving the canonical distribution of the non-central rectangular coordinates directly from the probability law for random samples from a p-variate normal population for the cases (i) one non-zero root and (ii) two non-zero roots for general p and (b) using this to obtain the canonical non-central Wishart distribution obtained by T. W. Anderson and M. A. Girshick for the same cases.

18. Confidence Interval Estimation for the Parameters of a Rectangular and an Exponential Population in Terms of Complete or Censored Samples. (Preliminary Report.) (By Title.) S. N. ROY and A. E. SARHAN, University of North Carolina.

In previous papers by the second author point estimation in the above situations was discussed. In the present paper, using the techniques given in previous papers by the first author, confidence intervals are given for parameters and certain statistically important

parametric functions in the case of rectangular and exponential populations (both general and special forms) in terms of complete samples. A method is also indicated of generalizing to the case of censored samples and to certain other populations as well.

19. Some Generalizations of Analysis of Variance and Covariance to the Case of Discrete Variates or of Grouping in Qualitative Categories. (By Title.) S. N. ROY and MARVIN KASTENBAUM.

Associated with any design there is a general cell—say an m-dimensional cell—in which we have a number of observations classified into, say, p-dimensional cells where p is the number of "variates" or "number of ways of classification." The whole data can now be regarded as being arranged in an $(m + p)$-way classification such that the m-cells and p-cells are, as it were, two different kinds of marginals. Bearing in mind the nature of this difference, the usual estimation and testing procedures in analysis of variance and covariance are generalized to the situations indicated above. The generalization of multivariate analysis of variance of means to the above situations will be discussed in a later paper.

20. Some Analytic Properties of Markoff Functions: Denumerable Case. (By Title.) DONALD G. AUSTIN, Syracuse University.

Let $p_{ij}(t)$, $0 < t < \infty$, $i, j = 1, 2, \cdots$, be the stationary transition matrix of a Markov chain. The author extends his earlier result (to appear in *Proc. Nat. Acad. Sci.*) that $Dp_{ii}(0) = -q_i > -\infty$ implies $p_{ij}(t)$ has a continuous derivative on $[0, \infty]$. It is shown that if $q_j < \infty$, $p_{ij}(t)$ has a continuous derivative on $[0, \infty]$. In either case $\lim_{t\to\infty} Dp_{ij}(t)$ exists and is equal to 0. If $q_i < \infty$, then $Dp_{ij}(t + s) = \Sigma_k\, Dp_{ik}(t)p_{kj}(s)$; if q_i , $q_j < \infty$, then $Dp_{ij}(t + s) = \Sigma_k\, p_{ik}(t)\, Dp_{kj}(t)$ for $t, s > 0$.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Thomas L. Austin, Jr., formerly Research Assistant at the University of Georgia, has accepted a position as Mathematician with the Dept. of Defense in Washington.

Professor Francisco Azorín P. of the University of Madrid, Spain, spent the academic year 1954–55 at the Universidad Central de Venezuela under UNESCO's Technical Assistance Program.

Maurice H. Belz has been appointed as Professor of Statistics in the University of Melbourne.

Francesco Bignardi was designated Statistical Fellow at the University of Palermo (Italy) where he is charged with the teaching of Social Statistics in the Statistical School. He is also chief of the Economical and Statistical Service of the Banco di Sicilia's Presidence.

Arthur B. Brown has been promoted to Professor of Mathematics at Queens College, Flushing, N. Y.

Dr. E. J. Gumbel was Visiting Professor for Mathematical Statistics at the Free University of Berlin (West) for the Summer Term 1955.

Harry M. Hughes, formerly Assistant Professor, University of California, Berkeley, is now Analytical Statistician, Dept. of Biometrics, USAF School of Aviation Medicine, Randolph Field, Texas.

Dr. H. Paul Kelley is now on active duty as a Research Psychologist at the Naval School of Aviation Medicine, U. S. Naval Air Station, Pensacola, Florida. He was formerly an Educational Testing Service Psychometric Fellow and more recently was with the U.S. Air Force Personnel and Training Research Center at Lackland Air Force Base, San Antonio, Texas.

Wharton F. Keppler has recently transferred from his position as Statistician, U.S. Naval Ordnance Test Station, Inyokern, California, to Elgin Air Force Base, Florida, where he is an Operations Analyst, Office of Operations Analysis, Dept. of Chief of Staff/Operations (DCS/O), Hq. Building, Air Proving Ground Command.

Robert J. Nichol is Statistician, Quality Control Group, Planning Dept., R.C.A. Service Co., Inc., Missile Test Project, Patrick Air Force Base, Florida.

Gottfried E. Noether of Boston University has been promoted to Associate Professor.

Don C. Price is now employed in the Engineering Department of the Goodyear Aircraft Corporation, Akron, Ohio.

Ronald Pyke having received his M.Sc. degree in Mathematical Statistics is continuing at the University of Washington as Research Assistant while working toward a Doctorate.

Robert L. Rogers recently resigned from his statistical and accounting duties with Stokely-Van Camp, Inc. in order to accept a position as Mathematician in the Computing Bureau with International Business Machines, Inc., Los Angeles, California.

C. H. Springer has recently accepted a position with the Aircraft Gas Turbine Development Department, General Electric Company, Cincinnati, Ohio. As Component Testing Evaluation Engineer, he will be engaged in the application of statistical principles to the design and evaluation of Jet Engine Research Testing Operations.

Jerome R. Steen has become Manager of Quality Control for the Radio and Television Division of Sylvania Electric Products Inc. He is at present located at Batavia, N. Y.

Dr. Joseph V. Talacko, Assistant Professor of Mathematics, Marquette University, Milwaukee, Wisconsin, returned in February from Berkeley to Milwaukee. He plans to spend the second half of his 1954/55 Ford Foundation Fellowship at the University of Chicago.

Dr. G. S. Watson has resigned from the Department of Statistics, University of Melbourne, to take up a new position as Senior Fellow, Dept. of Statistics, The Australian National University, Canberra.

William Wolman, formerly with Naval Inspector of Ordnance, Eastman Kodak Co., Rochester, N. Y., is now head of the Statistical Methodology and Reliability Section, Statistics Branch, Quality Control Division, Bureau of Ordnance, Navy Department in Washington, D. C.

James K. Yarnold, formerly a graduate student and Research Assistant at the Statistical Laboratory, University of California, Berkeley, is now a graduate student in Mathematical Statistics at the University of Illinois and a Research Assistant in the University of Illinois Training Research Laboratory.

Mr. H. Zindler is now Referent für Mathematische Statistik in Abteilung VIII des Statistischen Bundesamtes, Wiesbaden-Biebrich, Rheinstr. 25, Germany.

Educational Testing Service

The Educational Testing Service is offering for 1956–57 its ninth series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of $2,500 a year and are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School.

Suitable undergraduate preparation may consist either of a major in psychology with supporting work in mathematics, or a major in mathematics together with some work in psychology. However, in choosing fellows, primary emphasis

is given to superior scholastic attainment and demonstrated research ability rather than to specific course preparation. The closing date for completing applications is January 12, 1956. Information and application blanks will be available about October 1 and may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

New Members

The following persons have been elected to membership in the Institute
February 9, 1955 to May 11, 1955

Baumann, Carl O., B.A. (American International College), Assistant Statistician, Monsanto Chemical Company, Indian Orchard, Massachusetts, *221 Britton Street, Fairview Massachusetts.*

Beckwith, Richard E., B.S. (Stanford Univ.), Research Assistant, Case Institute of Technology, Cleveland 6, Ohio.

Birch, John J., B.S., (Brown Univ.), Graduate Student in Statistics, University of California, Berkeley, California, *540 Alcatraz Avenue, Oakland 9, California.*

Block, Aaron, B.A. (Brooklyn College), Analyst, City Planning Department, Office of Master Planning, 15 Park Row, New York, New York, *102-35 64th Road, Forest Hills 75, New York.*

Blum, Joseph, M.A. (George Washington Univ.), Machine Methods Analyst, Assistant Branch Chief of Spec. Processing Branch, National Security Agency, Washington 25, D. C., *4314 N. Pershing Drive, Arlington 3, Virginia.*

Boyd, Evelyn, Ph.D. (Yale Univ.), Mathematician, Department of the Army, Diamond Ordnance Fuse Laboratory, Washington 25, D. C., *1353 Ritchie Place, N. E., Washington 17, D. C.*

Breakwell, John Valentine, Ph.D. (Harvard Univ.), Senior Research Engineer, North American Aviation, Inc., Downey, California, *21 Temple Avenue, Long Beach 3, California.*

DeMarr, Ralph A., M.A. (Washington State College), Member of Technical Staff, Bell Telephone Laboratory, Whippany, New Jersey, *29 DeBary Place, Summit, New Jersey*

Dubay, Joseph A., M.A. (Harvard Univ.), Graduate Student, Committee on Statistics, University of Chicago, Chicago 37, Illinois, *22 Snell Hall, 5709 South Ellis Avenue, Chicago 37, Illinois.*

Friedman, Henry D., Ph.D. (Penn. State Univ.), Mathematician, General Electric Co., Electronics Park, Syracuse, New York, *Room 235, Bldg. 3.*

Guthrie, Donald Jr., B.Sc. (Stanford Univ.), Student and Department Teaching Assistant, Dept. of Mathematical Statistics, Columbia University, New York 27, New York.

Heit, Paul B., B.S. (C.C.N.Y.), Graduate Student, Columbia University, New York 27, New York, *1540 Walton Avenue, Bronx 52, New York.*

Hopkins, John W., Ph.D. (Univ. of London), Biometrician, Division of Applied Biology, National Research Council, Ottawa 2, Canada.

Inman, Patricia, B.A. (Reed College), Student and Research Assistant, Department of Mathematics, University of Oregon, Eugene, Oregon.

Kastenbaum, Marvin A., M.S. (North Carolina State College), Graduate Assistant, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina.

Kendall, G. R., M.A. (Univ. of Toronto), Meteorologist and Climatologist, Meteorological Division, Department of Transport, 315 Bloor Street W., Toronto, Ontario, Canada, *2099 Stovebank Road, R.R. #1, Port Credit, Ontario, Canada.*

Lanahan, James F., M.S. (Univ. of Michigan), Instructor of Mathematics, University of Detroit, Detroit 21, Michigan.

Madansky, Albert, A.B. (Univ. of Chicago), Student in Committee on Statistics, University of Chicago, Eckhart Hall, Chicago, Illinois, *1516 S. Kostner, Chicago 23, Illinois.*

Oneson, Thomas M., M.A. (Univ. of Illinois), Statistician, Owens-Corning Fiberglas Corp., Case Avenue, Newark, Ohio.

Paul, William H., B.S. (Penn. State College), Project Supervisor, Aircraft Instrumentation Display Studies, Stavid Engineering, Inc., 312 Park Avenue, Plainfield, New Jersey.

Perrin, Edward B., B.A. (Middlebury College), Graduate Student, Department of Mathematical Statistics, Columbia University, *179 S. Main Street, Barre, Vermont.*

Rothman, Stanley, M.A. (Columbia Univ.) Staff Member, RAMO Wooldridge Co., 8820 Bellanca Avenue, Los Angeles, California.

Sax, Edward, B.A. (Wayne Univ.), Analytical Statistician, *1121 New Hampshire Avenue, #403, Washington 6, D. C.*

Seibel, Melvin H., B.S. (Univ. of Illinois), Physicist, Statistical Section, Engineering and Technical Department, Army Electronic Proving Ground, Fort Huachuca, Arizona.

Stevens, Kerry N., B.A. (Central Wash. College), Graduate Student, University of Washington, Seattle 5, Washington, *120 East 110th, Seattle 55, Washington.*

Stuart, Alan, B.Sc. (Univ. of London), Senior Research Officer and Lecturer, Division of Research Techniques, London School of Economics, Houghton Street, Aldwych, London W.C. 2, England.

Suurballe, John W., M.Sc. (State Univ. of Iowa), Applied Mathematician, Farnsworth Electronics Co., Fort Wayne, Indiana, *2407 North Anthony, Fort Wayne, Indiana.*

Thampuran, D. V., M.S. (Univ. of Travancore), Research Assistant, Central Marine Fisheries Research Station, Madapam Camp., P.O., S. Rly, India.

Wartmann, Rolf, Dr. (Technische Hochschule Darmstadt), Sachbearbeiter für Technische Statistik, Verein Deutscher Eisenhüttenleut, Düsseldorf, Breite Str. 27, Germany.

Weibull, Martin, Fil. Dr. (The Royal Univ. of Lund), Docent, University of Lund, Lund, Sweden, Department of Statistics, Solvegatan 5, Lund, Sweden, *Revingeg 17, Lund, Sweden.*

Wendel, J. G., Ph.D., (California Inst. of Tech.), Associate Professor of Mathematics, Louisana State University, Baton Rouge, Louisana.

Wetzel, Wolfgang, Dr. rer Pol. (Freie Univ. Berlin), Wissenschaftlicher Assistant, Seminar für Statistik der Freie Universität Berlin, Berlin-Dahlem, Bachstelzenweg 29/31, *Berner Str. 19, Berlin—Lichterfelde West.*

REPORT OF THE CHAPEL HILL MEETING OF THE INSTITUTE

The 1955 Eastern Regional Meeting, sixty-fifth meeting of the Institute of Mathematical Statistics, was held in Chapel Hill, North Carolina, April 22–23, 1955. A meeting of the Biometric Society (Eastern North American Region) was held in Chapel Hill at the same time.

The following 80 members of the Institute registered for the meeting:

R. L. Anderson, T. W. Anderson, Helen Bozivich, R. H. Brunelle, Benjamin Buchbinder, D. L. Burkholder, J. M. Cameron, R. L. Carter, J. T. Chu, W. H. Clatworthy, A. C. Cohen, Jr., Theodore Colton, W. C. Connor, L. M. Court, E. L. Cox, Gertrude Cox, P. P. Crump, Claude de Courval, Cyrus Derman, Alfred Descloux, Elizabeth Doan, D. B. Duncan, Churchill Eisenhart, Lillian Elveback, Benjamin Epstein, S. M. Free, J. E. Freund, Seymour Geisser, H. S. Graf, B. G. Greenberg, S. W. Greenhouse, F. E. Grubbs, S. S. Gupta, R. J. Hader, Max Halperin, Boyd Harshbarger, Wassily Hoeffding, Jacob Horowitz, D. G. Horvitz, Harold Hotelling, W. G. Howe, J. S. Hunter, D. C. Hurst, Mohammad Iqbal, K. Ito, A. W. Kimball, Julius Lieblein, D. V. Lindley, Eugene Lukacs, J. H. MacKay, F. S. McFeely, H. A. Meyer, D. F. Morrison, M. E. Muller, V. N. Murty, C. R. Newell, G. E. Nicholson, Jr., G. E. Noether, Wyman Richardson, D. L. Richter, David Rosenblatt, Joan R. Rosenblatt, A. E. Sarhan, Roberto Sasso, F. E. Satterthwaite, M. A. Schneiderman, H. Smith, W. L. Smith, G. W. Snedecor, P. N. Somerville, Jacques St-Pierre, H. C. Sweeny, Z. Szatrowski, W. A. Thompson, Jr., M. E. Turner, Jr., M. C. K. Tweedie, Lionel Weiss, J. W. Wilkinson, R. L. Wine, Marvin Zelen.

The program was as follows:

FRIDAY, APRIL 22, 1955

8:30 a.m. Joint Session with Biometric Society.

Chairman: H. FAIRFIELD SMITH, North Carolina State College.

Papers:

1. *Life Testing in the Discrete Case*, FRANKLIN S. MCFEELY and JOHN E. FREUND, Virginia Polytechnic Institute.
2. *The Components of Variance and the Correlation Between Relatives in Symmetrical Random Mating Populations*, TED HORNER, Iowa State College.
3. *Tests of Hypotheses when the Decision is Based on Several Criteria (Preliminary Report)*, IRWIN MILLER and JOHN E. FREUND, Virginia Polytechnic Institute.
4. *Power Function of Procedures for Some Components of Variance Models*, HELEN BOZIVICH, Iowa State College.
5. *Preference Patterns for Decisions on Means*, R. LOWELL WINE and JOHN E. FREUND, Virginia Polytechnic Institute.

(Papers 1, 3, and 5 were on work sponsored by the Office of Ordnance Research, U.S. Army.)

10:30 a.m. Problems of Probability.

Chairman: EUGENE LUKACS, Office of Naval Research.

Papers:

1. *Differentiation of Markov Transition Functions*, DONALD G. AUSTIN, Syracuse University.
2. *An Extension of the Kolmogorov Limit Theorem*, JERRY BLACKMAN, Syracuse University.
3. *Non-recurrent Random Walks*, CYRUS DERMAN, Syracuse University.

2:00 p.m. Multivariate Analysis.

Chairman: HAROLD HOTELLING, University of North Carolina.

Papers:
1. *Principal Components and Factor Analysis*, T. W. ANDERSON, Columbia University.
2. *Some Contributions to Factor Analysis*, WILLIAM G. HOWE, Oak Ridge National Laboratory and University of North Carolina.
3. *Analysis of Variance of Correlated Variates with Heterogeneous Variances*, H. C. SWEENEY, Virginia Polytechnic Institute.

4:00 p.m. Contributed Papers I.

Chairman: GEORGE E. NICHOLSON, JR., University of North Carolina.

Papers:
1. *Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples. Part I. The Normal Distribution up to Samples of Size* 10, A. E. SARHAN and B. G. GREENBERG, University of North Carolina.
2. *An Application of Chung's Lemma to the Kiefer-Wolfowitz Stochastic Approximation Procedure*, CYRUS DERMAN, Syracuse University.
3. *Simplified Estimators Based on Order Statistics (Preliminary Report)*, BENJAMIN EPSTEIN, Wayne University.
4. *Distribution of the Difference Between the Two Largest Sample Values (Preliminary Report)*, A. ZINGER and J. ST-PIERRE, University of Montreal.
5. *Some Continuous Monte Carlo Methods for the Dirichlet Problem*, MERVIN E. MULLER, Cornell University.
6. *On the Distribution of the Number of Successes in Independent Trials*, WASSILY HOEFFDING, University of North Carolina.
7. *On the Solution of Truncated and Censored Sample Estimating Equations for Normal Populations*, A. C. COHEN, JR., University of Georgia.
8. *The Modified Mean Square Successive Difference and Related Statistics*, SEYMOUR GEISSER, University of North Carolina.
9. *The Distribution of the Ratios of Certain Quadratic Forms in Time Series* (By Title), SEYMOUR GEISSER, University of North Carolina.

SATURDAY, APRIL 23, 1955

8:30 a.m. Symposium on Relation Between Smoking and Mortality from Lung Cancer. (Co-sponsored by Biometric Society.)

Chairman: B. G. Greenberg, University of North Carolina.

Papers:
1. *Current Status of the Problem*, JEROME CORNFIELD, National Institute of Health.
2. *Needed Future Work*, WILLIAM HAENSZEL, National Cancer Institute.

Discussants: BOYD HARSHBARGER, Virginia Polytechnic Institute, DANIEL HORN, American Cancer Society.

2:00 p.m. Contributed Papers II.

Chairman: LIONEL WEISS, University of Virginia.

Papers:
1. *The "Inefficiency" of the Sample Median for Many Familiar Symmetric Distributions*, J. T. CHU, University of North Carolina.
2. *On Some Stochastic Models of Behavioral Interaction of Organization Theory*, DAVID ROSENBLATT, American University.

3. *On Inverting a Class of Patterned Matrices, Part* I, S. N. ROY and A. E. SARHAN, University of North Carolina.
4. *Convergence Properties of a General Stochastic Approximation Process* (*Preliminary Report*), DONALD BURKHOLDER, University of North Carolina.
5. *Distribution of Rounding Off Errors in Some Numerical Processes, Part* I, A. E. SARHAN, University of North Carolina.
6. *On a Measure of the Information Provided by an Experiment* (*Preliminary Report*), D. V. LINDLEY, University of Cambridge and University of Chicago.
7. *A Comparison Between Alternative Techniques Using Supplementary Information in Sample Survey Design* (*Preliminary Report*), EL MAHDY SAID, North Carolina State College (introduced by R. L. Anderson).
8. *The Canonical Distribution of the Non-central Rectangular Co-ordinates*, MISS ALEYAMMA GEORGE, University of North Carolina and University of Travancore (introduced by H. Hotelling).
9. *Confidence Interval Estimation for the Parameters of a Rectangular and an Exponential Population in Terms of Complete or Censored Samples* (*Preliminary Report*) (By Title), S. N. ROY and A. E. SARHAN, University of North Carolina.
10. *Some Generalizations of Analysis of Variance and Covariance to the Case of Discrete Variates or of Grouping in Qualitative Categories* (By Title), S. N. ROY and MARVIN KASTENBAUM, University of North Carolina.
11. *Some Analytic Properties of Markoff Functions: Denumerable Case* (By Title), DONALD G. AUSTIN, Syracuse University (introduced by C. Derman).

LIONEL WEISS
Associate Secretary

PUBLICATIONS RECEIVED

Tables of Sines and Cosines for Radian Arguments, National Bureau of Standards, Applied Mathematics Series 43, U.S. Government Printing Office, Washington, D. C., 1955, 278 pp., $3.00.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

THE INST[illegible] STATISTI[illegible]

President:

Henry Sc[illegible] [illegible]ley, California

President-[illegible]:

[illegible] [illegible] California

Secretary:

G. E. [illegible] Carolina [illegible] North Carolina

Treasurer:

[illegible] B[illegible] [illegible] C[illegible]

Editor:

[illegible] August 1, 19[illegible] [illegible] 1700 Main Street, Santa Moni[illegible]

[illegible] purpose of the In[illegible] [illegible]courage the development, dissemination [illegible] [illegible]tistics. Membership dues [illegible] [illegible] OF MATHEMATICAL STATISTICS [illegible] [illegible] United [illegible] in Canada and [illegible] [illegible] There are rates for students [illegible] [illegible] Inquiries [illegible] membership in the In[illegible] [illegible] Secretary of the Insti[illegible]

ANN[illegible] MEETING [illegible] NEW [illegible]K

To be held in [illegible] [illegible] Association and others.

(Abstracts of [illegible] [illegible] days [illegible] the speaker [illegible] Ab- [illegible]